

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:16:18 ; Search time 49.45 Seconds  
(without alignments)  
40.806 Million cell updates/sec

Title: US-09-461-684-1  
Perfect score: 109  
Sequence: 1 CXXXXXXXXXXXXXXXXXXXX 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

PIR-71:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	100	91.7	215	2	nucleoporin p62 ho
2	100	91.7	380	2	hypothetical prote
3	100	91.7	517	2	hypothetical prote
4	93	85.3	166	2	hypothetical prote
5	90	82.6	483	2	hypothetical prote
6	87	79.8	107	2	protein p1504.29
7	84	77.1	441	2	acidic phosphoprot
8	81	74.3	4550	2	hypothetical prote
9	80	73.4	784	2	hypothetical prote
10	78	71.6	529	2	hypothetical prote
11	78	71.6	2506377	2	SAR DNA-binding pr
12	78	71.6	1560	2	proliferation pote
13	77	70.6	229	2	nuclear protein SR
14	77	70.6	266	2	protein F9L1.30
15	77	70.6	3724	2	hypothetical prote
16	76	69.7	511	2	nucleolar protein
17	76	69.7	812	2	R075.1 protein (C
18	75	68.8	474	2	centromere/microtu
19	75	68.8	504	2	hypothetical prote
20	74	67.9	191	2	hypothetical prote
21	74	67.9	198	2	probable membrane
22	74	67.9	414	2	hypothetical prote
23	74	67.9	483	2	centromere/microtu
24	73	67.0	212	2	related to protein k
25	72.5	66.5	683	2	hypothetical prote
26	72	66.1	404	2	hypothetical prote
27	72	66.1	560	2	glutamic acid-rich
28	72	66.1	678	2	hypothetical prote
29	72	66.1	1701	2	probable erythrocy

30	72	66.1	4981	2	T18489	hypothetical prote
31	71	65.1	683	2	T34103	hypothetical prote
32	71	65.1	686	1	A44842	CGMP-gated ion cha
33	71	65.1	690	2	A42161	CGMP-gated cation
34	70.5	64.7	163	2	T42696	hypothetical prote
35	70	64.2	222	2	T28919	hypothetical prote
36	70	64.2	233	2	S55165	hypothetical prote
37	70	64.2	231	2	T48617	hypothetical prote
38	70	64.2	451	2	T16418	hypothetical prote
39	70	64.2	497	2	T29814	hypothetical prote
40	70	64.2	508	2	E71620	hypothetical prote
41	70	64.2	688	2	B42161	CGMP-gated cation
42	70	64.2	2013	2	C71610	probable membrane
43	69.5	63.8	2231	2	S53416	SEN1 protein - yea
44	69	63.3	142	2	S54481	hypothetical prote
45	69	63.3	167	2	S38112	hypothetical prote

## ALIGNMENTS

RESULT 1  
152523  
nucleoporin p62 homolog - rat (fragment)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 27-Feb-1997 #sequence\_revision 27-Feb-1997 #text\_change 05-Nov-1999  
C:Accession: 152523  
R:Wang, Z.O.; Akmal, K.M.; Kim, K.H.  
Biol. Reprod. 51, 1022-1030, 1994.  
A:Title: An unusual nucleoporin-related messenger ribonucleic acid is present in the  
A:Reference number: 152523; M0ID:95151924  
A:Accession: 152523  
A>Status: Preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-215 <RES>  
A:Cross-references: GB:S75997; NID:913245; PID:NAB3384.1; PID:913246  
A:Experimental source: testis

Query Match Best Local Similarity 91.7%; Score 100; DB 2; Length 215;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21  
Db 35 KKKKKKKKKKKKKKKKKKK 54

RESULT 2  
T46395  
hypothetical protein DKFP4341120.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C:Accession: T46395  
R:Ottewill, B.; Obermayer, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, January 2000  
A:Reference number: Z23031  
A:Accession: T46395  
A>Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-380 <AAA>  
A:Cross-references: EMBL:AL137556  
A:Experimental source: adult testis; clone DKFP4341120  
C:Genetics:  
A>Note: DKFP4341120.1

Query Match Best Local Similarity 91.7%; Score 100; DB 2; Length 380;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 KKKKKKKKKKKKKKKKKKK 21  
|||||

```
Db 355 KKKKKKKKKKKKKKKKKKKKK 374
RESULT 3
T49173
hypothetical protein T20N10.250 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 08-Dec-2000
C:Accession: T49173
R:D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; I
submitted to the Protein Sequence Database, April 2000
A:Reference number: 225017
A:Accession: T49173
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-517 <DAN>
A:Cross-references: EMBL:AL353032; GSPDB:GN00061; ATSP:T20N10.250
A:Experimental source: cultivar Columbia; BAC clone T20N10
C:Genetics:
A:Gene: ATSP:T20N10.250
A:Map position: 3
A:Introns: 312/3: 359/3: 444/3
C:Superfamily: Arabidopsis thaliana hypothetical protein F17J16.30

Query Match 91.7%; Score 100; DB 2; Length 517;
Best Local Similarity 100.0%; Pred. No. 0.003;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 21
Db 444 KKKKKKKKKKKKKKKKKKKKK 463

RESULT 4
T18513
hypothetical protein C0845C - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T18513
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: 218935
A:Accession: T18513
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-166 <LAN>
A:Cross-references: EMBL:Z98551; PIDN:CAB11123.2
C:Genetics:
A:Map position: 3
A:Introns: 19/1
A:Note: C0845C

Query Match 85.3%; Score 93; DB 2; Length 166;
Best Local Similarity 90.0%; Pred. No. 0.0067;
Matches 18; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 21
Db 37 KKKKKKKKKKKKKKKKKKKKK 56

RESULT 5
F71619
hypothetical protein PF0235W - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: F71619
R:Gardner, M.J.; Tetelijn, H.; Carucci, D.J.; Cummings, L.M.; Araavind, L.; Koonin, E.V.;
Science 282: 1126-1137, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MUID:99021743
A:Accession: F71619
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-483 <GAR>
A:Cross-references: GB:AE001382; GB:AE001362; NID:93845130; PIDN:AACT1836.1; PID:938,
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PF0235W

Query Match 82.6%; Score 90; DB 2; Length 483;
Best Local Similarity 85.0%; Pred. No. 0.026;
Matches 17; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 21
Db 449 KKKKKKKKKKKKKKKKKKKKK 468

RESULT 6
C86477
protein F1504.29 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C86477
R:Ritoeogis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alor
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marzic
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talli
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C86477
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-107 <STO>
A:Cross-references: GB:AE005172; NID:98778346; PIDN:AAF9354.1; GSPDB:GN00141
C:Genetics:
A:Gene: F1504.29
A:Map position: 1

Query Match 79.8%; Score 87; DB 2; Length 107;
Best Local Similarity 94.4%; Pred. No. 0.019;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 19
Db 29 KKKKKKKKKKKKKKKKKKKKK 46

RESULT 7
A48455
acidic phosphoprotein PCMA1g - Plasmodium chabaudi
C:Species: Plasmodium chabaudi
C>Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998
C:Accession: A48455
R:Deleersnijder, W.; Prasomsitt, P.; Tunpradubkul, S.; Hendrix, D.; Hamers-Casterma
Mol. Biochem. Parasitol. 56, 59-68, 1992
A:Title: Structure of a Plasmodium chabaudi acidic phosphoprotein that is associated
A:Reference number: A48455; MUID:93116806
A:Accession: A48455
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-441 <DEL>
A:Cross-references: GB:M95789; NID:9160602; PID:9160603
A:Experimental source: IP-PC1/C
A:Note: sequence extracted from NCBI backbone (NCBIN:121415, NCBI:P:121416)
```



C:Keywords: phosphoprotein

Query Match 77.1%; Score 84; DB 2; Length 441;  
Best Local Similarity 80.0%; Pred. No. 0.089;  
Matches 16; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21  
|||||:|||||:|||||

Db 394 KKKKKKKKKKKKKKKKKKK 413

RESULT 8

T18440  
hypothetical protein C0425w - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000

C:Accession: T18440  
R:Lawson, D.; Bowman, S.; Barrell, B.

submitted to the EMBL Data Library, August 1997  
A:Reference number: Z18935

A:Accession: T18440  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
A:Residues: 1-4550 <LAW>

A:Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325396; PIDN:CAB1121.1

C:Genetics:  
A:Map position: 3  
A:Note: C0425w

Query Match 74.3%; Score 81; DB 2; Length 4550;  
Best Local Similarity 44.2%; Pred. No. 0.77;  
Matches 19; Conservative 2; Mismatches 0; Indels 22; Gaps 1;

OY 1 CRRK-----KKKKKKKKKKKKKK 21  
|||||

Db 132 CRRKNTFYNIKRYENEKYYQINNIKRRKKKKKKKKKK 174

RESULT 9

T18452  
hypothetical protein C0560c - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000

C:Accession: T18452  
R:Lawson, D.; Bowman, S.; Barrell, B.

submitted to the EMBL Data Library, November 1998  
A:Reference number: Z18937

A:Accession: T18452  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
A:Residues: 1-784 <LAW>

A:Cross-references: EMBL:AL008970; NID:e1407852; PID:e132545; PIDN:CA15594.1

C:Genetics:  
A:Map position: 3  
A:Note: C0560c

Query Match 73.4%; Score 80; DB 2; Length 784;  
Best Local Similarity 75.0%; Pred. No. 0.31;  
Matches 15; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21  
|||||:|||||:|||||

Db 360 EKKKKKKKKKKKKKKKKKK 379

RESULT 10

T50609  
hypothetical protein DKFZp761B2423.1 - human (fragment)

C:Species: Homo sapiens (man)  
C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000

C:Accession: T50609  
R:Boecker, H.; Boecker, M.; Brandt, P.; Mewes, H.W.; Weil, B.; Wiemann, S.

submitted to the Protein Sequence Database, June 2000  
A:Reference number: Z25143

A:Accession: T50609  
A:Status: preliminary

A:Molecule type: mRNA  
A:Residues: 1-529 <AAA>

A:Cross-references: EMBL:AJ359564  
A:Experimental source: adult amygdala; clone DKFZp761B2423

C:Genetics:  
A:Note: DKFZp761B2423.1

Query Match 71.6%; Score 78; DB 2; Length 529;  
Best Local Similarity 80.0%; Pred. No. 0.37;  
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21  
|||||:|||||:|||||

Db 464 KKKKKKKKKKKKKKKKKKK 483

RESULT 11

T06377  
SAR DNA-binding protein-1 - garden pea

C:Species: Pisum sativum (garden pea)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 02-Jun-2000

C:Accession: T06377  
R:Halton, D.; Gray, J.C.

submitted to the EMBL Data Library, April 1998  
A:Description: cDNA encoding a pea SAR DNA-binding protein that shows homology to nuc

A:Reference number: Z15637  
A:Accession: T06377

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA

A:Residues: 1-560 <HAT>  
A:Cross-references: EMBL:AF061862; NID:g3132695; PIDN:AA16330.1; PID:g3132696

C:Genetics:  
A:Gene: SARBP-1  
C:Superfamily: garden pea SAR DNA-binding protein

Query Match 71.6%; Score 78; DB 2; Length 560;  
Best Local Similarity 75.0%; Pred. No. 0.38;  
Matches 15; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21  
|||||:|||||:|||||

Db 463 KKKKKKKKKKKKKKKKKKK 482

RESULT 12

T42727  
proliferation potential-related protein - mouse

C:Species: Mus musculus (house mouse)  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 02-Sep-2000

C:Accession: T42727  
R:Mitte, M.M.; Scott, R.E.

submitted to the EMBL Data Library, November 1998  
A:Reference number: Z22246

A:Accession: T42727  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA  
A:Residues: 1-1560 <RT>

A:Cross-references: EMBL:U083913; NID:g3858884; PID:g3858885; PIDN:AA172432.1

A:Experimental source: strain Balb/C  
C:Genetics:  
A:Gene: P2P-R

C:Function:  
A:Description: involved in hnRNP association and Rb1 binding  
C:Superfamily: RING finger homology

F:57-107/Domain: RING finger homology <RRN>

Query Match 71.6%; Score 78; DB 2; Length 1560;  
Best Local Similarity 80.0%; Pred. No. 0.74;  
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21  
DB 1497 KKKKKKKKKKKKKKKKK 1516

## RESULT 13

JC7219  
nuclear protein SR-25 - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 21-Jul-2000  
C:Accession: JC7219  
R:Sasahara, K.; Yamaoka, T.; Moritani, M.; Tanaka, M.; Iwanana, H.; Yoshimoto, K.; Miyag  
Biochem. Biophys. Res. Commun. 269, 444-450, 2000  
A:Title: Molecular cloning and expression analysis of a putative nuclear protein, SR-25.  
A:Reference number: JC7219; MID:20175222  
A:Accession: JC7219  
A:Molecule type: mRNA  
A:Residues: 1-229 <SAS>  
A:Cross-references: DDBJ:AB035383; NID:g7619895; PIDN:BAA94743.1; PID:g7619896  
A:Experimental source: MINE cell line  
C:Comment: This protein is a highly hydrophilic nuclear protein with a serine-arginine  
A:Splicing factors.  
C:Keywords: nucleus; RNA processing

Query Match 70.6%; Score 77; DB 2; Length 229;  
Best Local Similarity 75.0%; Pred. No. 0.27;  
Matches 15; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21  
DB 95 KKKKKKKKKKKKKKKKKKK 114

## RESULT 14

A86288  
protein F9L1.30 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: A86288  
R:Thoclogis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewart, K.;  
ansen, N.E.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.W.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MID:21016719  
A:Accession: A86288  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-266 <STO>  
A:Cross-references: GB:AE005172; NID:g5103832; PIDN:AAD39662.1; GSPDB:GNC0141  
C:Genetics:  
A:Gene: F9L1.30  
A:Map position: 1

Query Match 70.6%; Score 77; DB 2; Length 266;  
Best Local Similarity 70.0%; Pred. No. 0.3;  
Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21  
DB 1497 KKKKKKKKKKKKKKKKK 1516

DB 231 KKKKKKKKKKKKKKKKK 250

## RESULT 15

hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
C:Accession: T18427  
R:Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, August 1997  
A:Reference number: 218935  
A:Accession: T18427  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-3724 <LAW>  
A:Cross-references: EMBL:298547; NID:e1325376; PID:e1325379; PIDN:CAB1104.1  
C:Genetics:  
A:Introns: 307/1; 1545/2  
A>Note: C0335c

Query Match 70.6%; Score 77; DB 2; Length 3724;  
Best Local Similarity 75.0%; Pred. No. 1.6;  
Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 CKKKKKKKKKKKKKKKKKK 20  
DB 2201 CKKKKKKKKKKKKKKKK 2220

Search completed: July 1, 2002, 06:28:08  
Job time: 710 sec

Mon Jul 1 13:54:45 2002

us-09-461-684-1.rpt

---

Page 5

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:20:23 ; Search time 25.37 Seconds

(without alignments)  
32.050 Million cell updates/sec

Title: US-09-461-684-1

Perfect score: 109

Sequence: 1 CKKKKKKKKKKKKKKKKKKK 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84	77.1	441	1 PHPA_PLACH	002752 plasmodium
2	76	69.7	511	1 NOP5_YEAST	012499 saccharomyc
3	75	68.8	474	1 CBF5_SCHPO	014607 schizosacch
4	75	68.8	504	1 SIK1_YEAST	012460 saccharomyc
5	75	68.8	2404	1 SON_MOUSE	093447 mus musculu
6	75	68.8	2404	1 SON_MOUSE	093447 mus musculu
7	74	67.9	414	1 Y694_METJA	018583 homo sapien
8	74	67.9	474	1 CBF5_KLULA	058105 methanococc
9	74	67.9	483	1 CBF5_YEAST	013473 kluyveromyc
10	74	67.9	726	1 BRD3_HUMAN	033322 saccharomyc
11	72	66.1	479	1 CBF5_CANAL	015059 homo sapien
12	72	66.1	678	1 GARP_PLAFL	043101 candida alb
13	71	65.1	351	1 CG79_HUMAN	013816 plasmodium
14	71	65.1	686	1 CNG1_HUMAN	093388 homo sapien
15	70.5	64.7	1411	1 TCOR_HUMAN	013428 homo sapien
16	70	64.2	539	1 YJ08_YEAST	013428 homo sapien
17	70	64.2	599	1 HM21_HUMAN	094905 homo sapien
18	70	64.2	683	1 CNG1_RAT	047006 saccharomyc
19	70	64.2	684	1 CNG1_MOUSE	062927 r camp-gate
20	69.5	63.8	534	1 NOP5_RAT	029974 m camp-gate
21	69.5	63.8	2231	1 SENI_YEAST	099286 rattus norv
22	69	63.3	142	1 YMBH_YEAST	000416 saccharomyc
23	69	63.3	167	1 YK20_YEAST	003525 saccharomyc
24	69	63.3	723	1 SSRP_DROME	005344 drosophila
25	69	63.3	843	1 BLVR_BOVIN	003368 bos taurus
26	68.5	62.8	724	1 Y051_CAEEL	034600 caenorhabd
27	68	62.4	523	1 DBP3_YEAST	020447 saccharomyc
28	68	62.4	1178	1 MNNA_YEAST	036044 saccharomyc
29	67	61.5	118	1 Y093_CAEEL	010032 caenorhabd
30	67	61.5	690	1 CNG1_BOVIN	000194 b camp-gate
31	67	61.5	691	1 CNG1_CANEA	028279 b camp-gate
32	67	61.5	1002	1 IF2P_YEAST	039730 saccharomyc
33	67	61.5	1220	1 IF2E_HUMAN	060841 homo sapien

34	67	61.5	1362	1 BRD4_HUMAN	060885 homo sapien
35	66	60.6	481	1 CBF5_EMBU	043100 emeticella
36	66	60.6	487	1 CBF5_ASPTU	043102 aspergillus
37	66	60.6	687	1 Y001_SCHPO	013796 schizosacch
38	66	60.6	1153	1 A3D1_HUMAN	014617 homo sapien
39	66	60.6	1240	1 YN11_YEAST	033935 saccharomyc
40	65.5	60.1	508	1 NO60_DROME	044081 drosophila
41	65.5	60.1	514	1 DKC1_HUMAN	060883 homo sapien
42	65	59.6	217	1 K81_HYDAT	038978 hydra atten
43	65	59.6	271	1 Y55M_YEAST	033335 saccharomyc
44	65	59.6	320	1 YD33_YEAST	012117 saccharomyc
45	65	59.6	344	1 YR02_YEAST	038079 saccharomyc

## ALIGNMENTS

RESULT ID	1	PHPA_PLACH	STANDARD:	PRT:	441 AA.
AC	002752;				
DT	01-JUL-1993 (Rel. 26, Created)				
DT	01-JUL-1993 (Rel. 26, Last sequence update)				
DT	01-JUN-1994 (Rel. 29, Last annotation update)				
DE	Acidic phosphoprotein precursor (50 kDa antigen).				
GN	PCENAI.				
OS	Plasmodium chabaudi.				
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
OX	NCBI_TaxID=5825;				
RP	[1]				
RC	SEQUENCE FROM N.A.				
RA	STRAIN=IP-PC1.				
RA	MEDLINE=93116806; PubMed=1475002;				
RA	Deleersnijder W., Pracomisti P., Tungratadikul S., Hendrix D.,				
RA	Hammers-Castelman C., Hammers R.;				
RT	"Structure of a Plasmodium chabaudi acidic phosphoprotein that is				
RT	associated with the host erythrocyte membrane.";				
RT	Mol. Biochem. Parasitol. 56:59-68(1992).				
RL	FUNCTION: DURING INFECTION, THIS PHOSPHOPROTEIN PROBABLY MODULATES				
CC	THE STRUCTURE OF THE RED CELL MEMBRANE TO THE ADVANTAGE OF THE				
CC	PARASITE, ALTHOUGH ITS PRECISE FUNCTION IS NOT KNOWN.				
CC	-I- CYTOPLASMIC LOCATION: PERIPHERAL MEMBRANE PROTEIN ON THE				
CC	THROUGHOUT THE ENTIRE ERYTHROCYTE CYCLE.				
CC	-I- MISCELLANEOUS: ASSOCIATED WITH THE HOST RED CELL MEMBRANE				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	-----				
CC	EMBL: M95789; AAA29732.1; -				
DR	PIR: A48458; A48455.				
KW	Phosphorylation; Signal; Antigen; Membrane; Repeat; Erythrocyte.				
FT	SIGNAL	1	441	OR 24 (POTENTIAL).	
FT	DOMAIN	16	313	ACIDIC PHOSPHOPROTEIN.	
FT	REPEAT	186	193	16 X 8 AA TANDEM REPEATS.	
FT	REPEAT	186	193	1-1.	
FT	REPEAT	194	201	1-2.	
FT	REPEAT	202	209	1-3.	
FT	REPEAT	210	217	1-4.	
FT	REPEAT	218	225	1-5.	
FT	REPEAT	226	233	1-6.	
FT	REPEAT	234	241	1-7.	
FT	REPEAT	242	249	1-8.	
FT	REPEAT	250	257	1-9.	
FT	REPEAT	258	265	1-10.	
FT	REPEAT	266	273	1-11.	
FT	REPEAT	274	281	1-12.	
FT	REPEAT	282	289	1-13.	

FT REPEAT 290 297 1-14.  
FT REPEAT 298 305 1-15.  
FT REPEAT 306 313 1-16.  
FT DOMAIN 353 370 2 X 9 AA TANDEM REPEATS.  
FT REPEAT 353 360 2-1.  
FT REPEAT 361 368 2-2.  
FT DOMAIN 371 417 LYS-RICH (BASIC).  
FT CARBOHYD 21 21 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 441 AA; 49708 MW; DB85E83E795E7E5 CRC64;

Query Match 77.1%; Score 84; DB 1; Length 441.  
Best Local Similarity 80.0%; Pred No. 0.023;  
Matches 16; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21  
DB 394 KKKKKKKKKKKKKKKKK 413

RESULT 2  
NOP5\_YEAST STANDARD; PRT; 511 AA.  
ID NOP5\_YEAST  
AC O12499;  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Nucleolar protein NOP58 (Nucleolar protein NOP5).  
GN NOP58 OR NOP5 OR YOR310C OR O6108.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=9288C; FY1679;  
RA Pearson B.M., Hernandez Y., Wolf S.S., Kalogeropoulos A., Schweizer M.,  
RL Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.  
RN [1]  
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.  
RC STRAIN=W303;  
RX MEDLINE=98298165; PubMed=9632712;  
RA Wu P., Brockenhrough J.S., Metcalfe A.C., Chen S., Aris J.P.;  
RT pre-18S rRNA processing in yeast.  
RL J. Biol. Chem. 273:16453-16463(1998).  
CC -1- FUNCTION: REQUIRED FOR PRE-18S RNA PROCESSING. MAY BIND  
MICROTUBULES.  
CC -1- SUBUNIT: INTERACTS WITH NOP56 AND NOP1.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.  
CC -1- SIMILARITY: BELONGS TO THE NOP5/NOP56 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X90565; CAA62165.1; -  
CC EMBL; Z75217; CAA99630.1; -  
CC EMBL; AF056070; AAC39484.1; -  
CC SCD; S0005837; NOP58.  
DR InterPro: IPR002687; NOP.  
DR Pfam: PF01798; NOP. 1.  
DR ProDom: PD004104; NOP. 1.  
KW Ribosome biogenesis; Nuclear protein; rRNA processing.  
FT DOMAIN 441 511 ASP/GLU/LYS-RICH.  
SQ SEQUENCE 511 AA; 56956 MW; 8A2889448B2A13E2 CRC64;

Query Match 69.7%; Score 76; DB 1; Length 511;  
Best Local Similarity 70.0%; Pred. No. 0.15;  
Matches 14; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21  
DB 480 KKKKKKKKKKKKKKKKK 499

RESULT 3  
CBF5\_SCHPO STANDARD; PRT; 474 AA.  
ID CBF5\_SCHPO  
AC O14007;  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Centromere/microtubule binding protein cbf5 (centromere-binding factor  
DE 5) (Nucleolar protein cbf5).  
GN CBF5 OR SPAC29A4.04C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA Brown D., Churcher C.M., Wood V., Barrell B.G., Rajandream M.A.;  
RL Submitted (Jul-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A  
CC CENTROMERE DNA-CBF5-BINDING FACTOR AND IS INVOLVED IN MITOTIC  
CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH (BY  
CC SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z97210; CAP10131.1; -  
CC InterPro: IPR002478; PUA.  
DR InterPro: IPR002501; Trub.N.  
DR Pfam: PF01472; PUA. 1.  
DR Pfam: PF01509; Trub.N. 1.  
DR SMART: SMO0359; PUA. 1.  
KW Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.  
FT DOMAIN 434 468 7 X 3 AA APPROXIMATE TANDEM REPEATS OF  
K-K-E.  
FT REPEAT 443 445 1.  
FT REPEAT 450 452 2.  
FT REPEAT 454 456 3.  
FT REPEAT 457 459 4.  
FT REPEAT 460 462 5.  
FT REPEAT 463 465 6.  
FT REPEAT 466 468 7.  
SQ SEQUENCE 474 AA; 53110 MW; B8C9896C5FAEB71 CRC64;

Query Match 68.8%; Score 75; DB 1; Length 474;  
Best Local Similarity 73.7%; Pred. No. 0.18;  
Matches 14; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 20  
DB 454 KKKKKKKKKKKKKKKKK 472

RESULT 4

SIR1\_YEAST  
ID SIR1\_YEAST STANDARD: PRT: 504 AA.  
AC 012460:  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE SIR1 protein (Nucleolar protein NOP56).  
OS SIR1 OR NOP56 OR YLR197W OR I8167.9.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota: Fungi: Ascomycota: Saccharomycotina: Saccharomycetes:  
OC Saccharomycetales: Saccharomycetaceae: Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=YPH1;  
RX MEDLINE=96040178; PubMed=7547500;  
RA Morin P.J., Downs J.A., Snodgrass A.M., Gilmore T.D.;  
RT "Genetic analysis of growth inhibition by GAL4-L Kappa B-alpha in  
RT Saccharomyces cerevisiae";  
RL Cell Growth Differ. 6:789-798(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=5288C / AB972;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
RA Favello A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,  
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,  
RA Johnston L., Langston Y., Latreille P., Mardis E., Menezes S.,  
RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,  
RA Taich S., Trevasakis E., Vaudin M., Vignati D., Wilcox L., Wilson R.,  
RA Wohldman P., Waterston R.;  
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP CHARACTERIZATION, AND MUTAGENESIS.  
RX MEDLINE=98038777; PubMed=9372940;  
RA Gautier T., Berges T., Tollervey D., Hurt E.;  
RT "Nucleolar KEX2 repeat proteins Nop56p and Nop58p interact with Nop1p  
RT and are required for ribosome biogenesis";  
RL Mol. Cell. Biol. 17:7088-7098(1997).  
CC -1- FUNCTION: REQUIRED FOR 60S RIBOSOMAL SUBUNIT SYNTHESIS.  
CC -1- SUBUNIT: INTERACTS WITH NOP1 AND NOP58.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.  
CC -1- SIMILARITY: BELONGS TO THE NOP5/NOP56 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U0237; AAC49066.1; -  
DR EMBL: U14913; AAB67431.1; -  
DR SGD: S0004187; SIR1.  
DR InterPro: IPR002687; NOP.  
DR Pfam: PF01798; NOP; 1.  
DR ProDom: PD004104; NOP; 1.  
KM Ribosome biogenesis; Nuclear  
FT DOMAIN 443 504  
FT MOTAEN 333 333  
FT V->A: REDUCED GROWTH RATE AT ALL  
FT TEMPERATURES: WHEN ASSOCIATED WITH R-385.  
FT Y->C: AT 37 DEGREES, GROWTH SLOWS AFTER 6  
FT MUTAGEN 355 355  
FT TO 8 HOURS AND CELL DIVISION STOPS AFTER  
FT 20 HOURS.  
FT M->R: REDUCED GROWTH RATE AT ALL  
FT TEMPERATURES: WHEN ASSOCIATED WITH A-333.  
FT MUTAGEN 385 385  
FT M->R: REDUCED GROWTH RATE AT ALL  
FT TEMPERATURES: WHEN ASSOCIATED WITH A-333.  
SQ SEQUENCE 504 AA: 56864 KM: F8522A5870EFA842 CRC64:  
Query Match 68.8%; Score 75; DB 1; Length 504;  
Best Local Similarity 70.0%; Pred. No. 0.19;  
Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 21  
DB 465 KKKKKKKKKKKKKKKKKKK 484  
RESULT 5  
ID SON\_MOUSE STANDARD: PRT: 2404 AA.  
AC 090X47; 090XP5; 090C06; 090C12;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE SON protein.  
GN SON.  
OS Mus musculus (Mouse).  
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:  
OC Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RC STRAIN=129/Sv;  
RX MEDLINE=20408886; PubMed=10950926;  
RA Wynn S.L., Fisher R.A., Pagel C., Price M., Liu Q.Y., Khan I.M.,  
RA Zammit P., Dedarri K., Mazzanti W., Kessling A., Lee J.S., Bulawa L.,  
RT "Organization and conservation of the GAR1/SON/DONSON locus in mouse  
RT and human genomes";  
RL Genomics 68:57-62(2000).  
RN [2]  
RP SEQUENCE OF 1-116 FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Hippocampus, Small intestine, and Tongue;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Akawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
RA Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,  
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
RA Kuentz P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schmitt L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Funuo M., Bono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Butt C., Fletcher C., Fujita K., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wyshak-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohetsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection";  
RL Nature 409:685-690(2001).  
CC -1- FUNCTION: Transcriptional repressor. Binds to the consensus DNA  
CC sequence: 5'-GAGT[AN(CG)]AG[CC-3'. Might protect cells from  
CC apoptosis. Might be involved in pre-mRNA splicing (By similarity).  
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are  
CC produced by alternative splicing.  
CC -1- TISSUE SPECIFICITY: Widely expressed.  
CC -1- DOMAIN: Contains 8 types of repeats which are distributed in 3  
CC regions.  
CC -1- SIMILARITY: CONTAINS 1 G-PATCH DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 DBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF193506; AAF23120.1; -  
DR EMBL: AF193595; AAF23120.1; JOINED.

DR	EMBL	AF1935596	AAF23120.1	JOINED	13 X 10 AA TANDEM REPEATS OF L-A [ST]-
DR	EMBL	AF1935597	AAF23120.1	JOINED	[MSG-[TS]-MDSM.
DR	EMBL	AF1935598	AAF23120.1	JOINED	11 X 7 AA TANDEM REPEATS OF [DR]-P-Y-R-
DR	EMBL	AF1935599	AAF23120.1	JOINED	[LI][AG][OHP].
DR	EMBL	AF1936000	AAF23120.1	JOINED	14 X 6 AA REPEATS OF [ED]-R-S-M-M-S.
DR	EMBL	AF193601	AAF23120.1	JOINED	1-1.
DR	EMBL	AF193602	AAF23120.1	JOINED	1-2.
DR	EMBL	AF193603	AAF23120.1	JOINED	1-3.
DR	EMBL	AF193604	AAF23120.1	JOINED	1-4.
DR	EMBL	AF193605	AAF23120.1	JOINED	1-5.
DR	EMBL	AF193607	AAF23121.1	-	1-6.
DR	EMBL	AK019312	BAB31658.1	-	1-7.
DR	EMBL	AK019081	BAB31536.1	-	1-8.
DR	EMBL	AK008478	BAB25691.1	-	1-9.
DR	EMBL	AK008256	BAB25562.1	-	1-10.
DR	MGD	MGI:98353	Son.	-	1-11.
DR	InterPro	IPR000159	DS_RBD.	-	1-12.
DR	InterPro	IPR000467	G_patch.	-	1-13.
DR	Pfam	PF000035	darm.1	-	1-14.
DR	Pfam	PF01585	G_patch.1	-	3 X 11 AA TANDEM REPEATS OF P-P-L-P-P-E-E-
DR	SMART	SM00443	G_patch.1	-	P-P-[TWE]-[MTC].
DR	PROSITE	PS50137	DS_RBD.1	-	7 X 7 AA REPEATS OF P-S-R-R-S-R-[TS].
KW	RNA-binding	DNA-binding	Nuclear protein	Repeat	1-1.
KW	Alternative splicing				1-2.
FT	DOMAIN	721	850		1-3.
FT	DOMAIN	867	943		1-4.
FT	DOMAIN	961	1080		1-5.
FT	REPEAT	961	966		1-6.
FT	REPEAT	969	974		1-7.
FT	REPEAT	976	981		1-8.
FT	REPEAT	985	990		1-9.
FT	REPEAT	993	998		1-10.
FT	REPEAT	1001	1006		1-11.
FT	REPEAT	1010	1015		1-12.
FT	REPEAT	1018	1023		1-13.
FT	REPEAT	1026	1031		1-14.
FT	REPEAT	1035	1040		3 X 11 AA TANDEM REPEATS OF P-P-L-P-P-E-E-
FT	REPEAT	1044	1049		P-P-[TWE]-[MTC].
FT	REPEAT	1055	1060		7 X 7 AA REPEATS OF P-S-R-R-S-R-[TS].
FT	REPEAT	1066	1071		1-1.
FT	REPEAT	1075	1080		1-2.
FT	DOMAIN	1101	1133		1-3.
FT	DOMAIN	1910	1979		1-4.
FT	REPEAT	1910	1916		1-5.
FT	REPEAT	1938	1944		1-6.
FT	REPEAT	1945	1951		1-7.
FT	REPEAT	1952	1958		1-8.
FT	REPEAT	1959	1965		1-9.
FT	REPEAT	1966	1972		1-10.
FT	REPEAT	1973	1979		1-11.
FT	DOMAIN	1919	1990		1-12.
FT	REPEAT	1919	1937		1-13.
FT	REPEAT	1980	1990		1-14.
FT	DOMAIN	1991	2017		3 X 11 AA TANDEM REPEATS OF P-S-R-R-S-R-[TS].
FT	DOMAIN	2283	2329		1-1.
FT	DOMAIN	2349	2404		1-2.
FT	VARSPPLIC	2086	2086		1-3.
FT	VARSPPLIC	2087	2404		1-4.
FT	SEQUENCE	2404	261428		1-5.

```

Query Match:      68.8%;  Score 75;  DB 1;  Length 2404;
Best Local Similarity 75.0%;  Pred. No. 0.6;
Matches 15;  Conservative 2;  Mismatches 3;  Indels 0;  Gaps 0;

OY 2 kkkkkkkkkkkkkkkkkk 21
    || || ||||| ||: ||

```

**RESULT**      **6**

ID	SON_HUMAN	STANDARD:	PRT: 2426 AA.
AC	P18583;	G95Y981; G9UPY0; Q14120; O14487; Q9UKP9; Q9H7B1; Q9P070;	
NC	Q9P072;		
DT	01-NOV-1990 (Rel. 16, Created)		
DT	01-MAR-2002 (Rel. 41, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	SON protein (SON3) (negative regulatory element-binding protein) (NRE binding protein) (DBP-5) (Box antagonist selected in saccharomyces 1) (BASSI) (protein C21orf50).		
DE	SON OR NREP OR DBP5 OR C21ORF50 OR KIAA1019.		
OS	Homo sapiens (Human).		
CC	Eukaryote; Metazoa; Choriata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A. (ISOFORM A; B; C; D; E AND F).		
RX	MEDLINE=21564202; PubMed=11707072;		
RA	Reymond A., Friedl M., Neegardt Henrichsen C., Chapot F.,		
RA	Deutsh S., Ueda C., Rossier C., Lyle R., Guttpont M.,		
RA	Anouarakis S.E.;		
RT	"From PRKDS and open reading frames to cDNA Isolation: revisiting the		
RT	Human Chromosome 21 Transcription Map.";		
RL	Genomics 78:46-54(2001).		
RN	[2]		
RP	SEQUENCE FROM N.A. (ISOFORM G).		
RC	TISSUE=Liver;		
RX	MEDLINE=21316479; PubMed=11306577;		
RA	Sun C.-T., Lo W.-Y., Wang I.-H., Lo Y.-H., Shiou S.-R., Lai C.-K.,		
RA	Ting L.-P.;		
RT	"Transcription repression of human hepatitis B virus genes by negative		
RT	regulatory element-binding protein/SON.";		
RL	J. Biol. Chem. 276:24059-24067(2001).		
RN	[3]		
RP	SEQUENCE OF 1-689 FROM N.A. (ISOFORM H).		
RC	TISSUE=Placenta;		
RA	Casadei R., Stroppoli P., D'Addabbo P., Canaider S., Lenzi L.,		
RA	Vitale L., Giannone S., Carinci P., Zannoni M.;		
RL	Submitted (OCT-2001) to the EMBL/Genbank/DDJ databases.		
RN	[4]		
RP	SEQUENCE OF 1-130 FROM N.A.		
RC	TISSUE=Smooth muscle;		
RA	Kawakami T., Noguchi S., Itch T., Shigeta K., Senba T., Matsumura K.,		
RA	Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujitara T., Ono T.,		
RA	Yanada K., Fujii Y., Ozaki K., Hiroo M., Ohmori Y., Ota T., Suzuki Y.,		
RA	Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,		
RA	Isegai T., Sugano S.;		
RT	"NDO human cDNA sequencing project.";		
RT	Submitted (Aug-2000) to the EMBL/Genbank/DDJ databases.		
RL	[5]		
RN	SEQUENCE OF 1-114 FROM N.A.		
RC	TISSUE=Blood;		
RA	Ye M., Zhang Q.H., Zhou J., Shen Y., Wu X.Y., Gan Z.Q., Wang L.,		
RA	Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;		
RT	"Human partial CDS from cd34+ stem cells."		
RL	Submitted (May-1999) to the EMBL/Genbank/DDJ databases.		
RN	[6]		
RP	SEQUENCE OF 437-2426 FROM N.A. (ISOFORM B).		
RC	TISSUE=Brain;		
RX	MEDLINE=99397452; PubMed=10470851;		
RA	Kikuno R., Nagase T., Ishikawa K.-I., Hirosewa M., Miyajima N.,		
RA	Tanaka A., Kotani H., Nomura N., Ohara O.;		
RT	"Prediction of the coding sequences of unidentified human genes. XIV.		
RT	The complete sequences of 100 new cDNA clones from brain which code		
RL	for large proteins in vitro.";		
RL	DNA Res. 6:197-205(1999).		
RN	[7]		
RP	SEQUENCE OF 554-2426 FROM N.A. (ISOFORM A).		
RX	MEDLINE=92049296; PubMed=1944235;		

RA Chumakov I.M., Berdichevskii F.B., Sokolova N.V., Reznikov M.V.,  
 RA Prasolov V.S.,  
 RT Identification of a protein product of a novel human gene SON and  
 RT the biological effect upon administering a changed form of this gene  
 RT into mammalian cells."  
 RL Mol. Biol. (Mosk) 25:731-740(1991).  
 RN [8]  
 RP SEQUENCE OF 709-1079 FROM N.A. (ISOFORM I).  
 RC TISSUE=Placenta;  
 RX MEDLINE=93062855; PubMed=1435774;  
 RA Bliskovskii V.V., Kirillov A.V., Zakhariev V.M., Chumakov I.M.;  
 RT "The human son gene: the large and small transcripts contains various  
 RT 5'-terminal sequences."  
 RL Mol. Biol. (Mosk) 26:807-812(1992).  
 RN [9]  
 RP SEQUENCE OF 1009-1131 FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=93062884; PubMed=1435773;  
 RA Bliskovskii V.V., Berdichevskii F.B., Tkachenko A.V., Belova M.E.,  
 RA Chumakov I.M.;  
 RT "Coding part of the son gene small transcript contains four areas of  
 RT complete tandem repeats."  
 RL Mol. Biol. (Mosk) 26:793-806(1992).  
 RN [10]  
 RP SEQUENCE OF 1145-2426 FROM N.A. (ISOFORM F).  
 RX MEDLINE=93048367; PubMed=1424986;  
 RA Mattioni T., Hume C.R., Konigorski S., Hayes P., Osterweil Z.,  
 RA Lee J.S.;  
 RT "A cDNA clone for a novel nuclear protein with DNA binding  
 RT activity."  
 RL Chromosoma 101:618-624(1992).  
 RN [11]  
 RP SEQUENCE OF 1692-2175 FROM N.A. (ISOFORM A).  
 RX MEDLINE=93039788; PubMed=3054499;  
 RA Berdichevskii F.B., Chumakov I.M., Kiselev L.L.;  
 RT "Decoding of the primary structure of the son3 region in human  
 RT genome: Identification of a new protein with unusual structure and  
 RT homology with DNA-binding proteins."  
 RL Mol. Biol. (Mosk) 22:794-801(1998).  
 RN [12]  
 RP SEQUENCE OF 1939-2426 FROM N.A. (ISOFORM J).  
 RC TISSUE=Cerebellum;  
 RX MEDLINE=99439804; PubMed=10509013;  
 RA Greenhalt W., Lee J., Chaudhuri B.;  
 RT "A selection system for human apoptosis inhibitors using yeast."  
 RL Yeast 15:1307-1321(1999).  
 CC -1- FUNCTION: Represses hepatitis B virus (HBV) core promoter activity  
 CC and transcription of HBV genes and production of HBV virions.  
 CC Binds to the consensus DNA sequence: 5'-GAGTAN[CG][AG]CC-3'.  
 CC Might protect cells from apoptosis. Might be involved in pre-mRNA  
 CC splicing (by similarity).  
 CC -1- SUBCELLULAR LOCATION: Nuclear with a speckled distribution.  
 CC -1- ALTERNATIVE PRODUCTS: 10 isoforms: A, B, C, D, E, F (shown here),  
 CC G, H, I and J; may be produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Widely expressed, with the higher expression  
 CC seen in leukocyte and heart.  
 CC -1- DOMAIN: Contains 8 types of repeats which are distributed in 3  
 CC regions.  
 CC -1- MISCELLANEOUS: Colocalizes with the pre-mRNA splicing factor  
 CC SF2/SC-35.  
 CC -1- SIMILARITY: CONTAINS 1 G-PATCH DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 DBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.  
 CC -1- CAUTION: ISOFORM A SEQUENCE FROM REF.7 DIFFERS FROM THAT SHOWN  
 CC DUE TO A FRAMESHIFT.  
 CC -1- CAUTION: ISOFORM F SEQUENCE FROM REF.10 DIFFERS FROM THAT SHOWN  
 CC DUE TO A FRAMESHIFT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch).  
 CC -----  
 DR EMBL; AF380179; AAL34497.1; -  
 DR EMBL; X63753; CAA45282.1; ALT\_FRAME.  
 DR EMBL; M36428; AAB36624.1; -  
 DR EMBL; AF380180; AAL34498.1; -  
 DR EMBL; AF380181; AAL34499.1; -  
 DR EMBL; AF380182; AAL34500.1; -  
 DR EMBL; AF380183; AAL34501.1; -  
 DR EMBL; AF380184; AAL34502.1; -  
 DR EMBL; AY026895; AAK07692.1; -  
 DR EMBL; AF435977; AAL30810.1; -  
 DR EMBL; X63751; CAC69885.1; -  
 DR EMBL; AB028942; BAA82971.1; -  
 DR EMBL; X63071; CAA44793.1; ALT\_FRAME.  
 DR EMBL; AF139897; AAD50078.1; -  
 DR EMBL; S47238; AAB23945.1; -  
 DR EMBL; AK024752; BAB14985.1; -  
 DR EMBL; AF161428; AAF28988.1; -  
 DR EMBL; AF161430; AAF28990.1; -  
 DR PIR; PNO099; PNO099.  
 DR MIM; 182465; -  
 DR InterPro; IPR001159; DS\_RBD.  
 DR InterPro; IPR000467; G\_patch.  
 DR Pfam; PF00035; derm; 1.  
 DR Pfam; PF01585; G\_patch; 1.  
 DR SMART; SM00358; DSRM; 1.  
 DR SMART; SM00443; G\_patch; 1.  
 DR PROSITE; PS50137; DS\_RBD; 1.  
 DR PROSITE; PS50174; G\_patch; 1.  
 DR RNA-binding; DNA-binding; Nuclear protein; Repeat;  
 KW Alternative splicing.  
 FT FT 726 895  
 FT FT 912 988  
 FT FT 1006 1126  
 FT FT 1006 1011  
 FT FT 1014 1019  
 FT FT 1021 1026  
 FT FT 1030 1035  
 FT FT 1038 1043  
 FT FT 1046 1051  
 FT FT 1055 1060  
 FT FT 1063 1068  
 FT FT 1071 1076  
 FT FT 1080 1085  
 FT FT 1089 1094  
 FT FT 1100 1105  
 FT FT 1111 1116  
 FT FT 1121 1126  
 FT FT 1147 1179  
 FT FT 1359 1390  
 FT FT 1925 1994  
 FT FT 1925 1931  
 FT FT 1953 1959  
 FT FT 1960 1966  
 Query Match 68.8%; Score 75; DS 1; Length 2426;  
 Best Local Similarity 75.0%; Pred. No. 0.51;  
 Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 KKKKKKKKKKKKKKKKKK 21  
 DB 109 KKKKKKKKKKKKKKKKKK 128  
 ID Y694\_METJA  
 AC Q58105;  
 RESULT 7  
 Y694\_METJA STANDARD; PRT; 414 AA.  
 ID Y694\_METJA  
 AC Q58105;



DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein MJ0694.  
 GN MJ0694.  
 OS Methanococcus jannaschli.  
 OC Archaea: Euryarchaeota: Methanococcales; Methanococcaceae;  
 CC Methanococcus  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RA MEDLINE=96337999; Pubmed=8688087;  
 RA Bilt C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.L.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
 RA Utermack T.R., Kelley J.M., Peterson J.D., Sadow P.M., Hanna M.C.,  
 RA Cotton H.P., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;  
 RA "complete genome sequence of the methanogenic archaeon, Methanococcus  
 RT jannaschli.";  
 RL Science 273:1058-1073(1996).  
 CC -1- SIMILARITY: BELONGS TO THE NOP5/NOP56 FAMILY.  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U67516; AAB98689.1; -  
 DR TIGR: MJ0694; -  
 DR InterPro: IPR002687; Nop.  
 DR Pfam: PF01798; Nop; 1.  
 DR ProDom: PD004104; Nop; 1.  
 KW Hypothetical protein; Complete proteome.  
 FT DOMAIN 349 414 ASP/GRU/LYS-RICH.  
 FT SEQUENCE 414 AA; 47799 MW; A9092EFC3C82C407 CRC64;  
 SQ

Query Match 67.9%; Score 74; DB 1; Length 414;  
 Best Local Similarity 75.0%; Pred. No. 0.2;  
 Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21  
 DB 377 KKKKKKKKKKKKKKKKK 396

RESULT 8  
 CBFS\_KLULA STANDARD; PRT; 474 AA.  
 AC 013473;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Centromere/microtubule binding protein CBFS (Centromere-binding factor  
 DE 5) (Nucleolar protein CBFS).  
 GN CBFS.  
 OS Kluyveromyces fragilis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 OX NCBI\_TaxID=28985;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JBD100;  
 RA Winkler A.A., Bobok A., Zonneveld B.J.M., Steensma H.Y.,  
 RA Hooykaas P.J.J.;

RT "The lysine-rich C-terminal repeats of the centromere-binding factor  
 RT 5 (Cbfs) of Kluyveromyces fragilis are not essential for function."  
 RL Yeast 14:37-48(1998).  
 CC -1- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A  
 CC CENTROMERE DNA-CBF3-BINDING FACTOR AND IS INVOLVED IN MITOTIC  
 CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH. MAY BE IN  
 CC SOME WAY ASSOCIATED WITH THE CBF3 110 KDA SUBUNIT (CBF3A) (BY  
 CC SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF008563; AAC64862.1; -  
 DR InterPro: IPR002478; PUA.  
 DR InterPro: IPR002501; Trub\_N.  
 DR Pfam: PF01472; PUA; 1.  
 DR SMART: SM00359; Trub\_N; 1.  
 KW Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.  
 FT DOMAIN 431 460  
 FT REPEAT 431 433 1.  
 FT REPEAT 434 436 2.  
 FT REPEAT 437 439 3.  
 FT REPEAT 440 442 4.  
 FT REPEAT 443 445 5.  
 FT REPEAT 446 448 6.  
 FT REPEAT 449 451 7.  
 FT REPEAT 452 454 8.  
 FT REPEAT 455 457 9.  
 FT SEQUENCE 474 AA; 53630 MW; 95306E7FEA756C CRC64;  
 SQ

Query Match 67.9%; Score 74; DB 1; Length 474;  
 Best Local Similarity 70.0%; Pred. No. 0.22;  
 Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21  
 DB 434 KKKKKKKKKKKKKKKKK 453

RESULT 9  
 CBFS\_YEAST STANDARD; PRT; 483 AA.  
 AC P33322;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Centromere/microtubule binding protein CBFS (centromere-binding factor  
 DE 5) (Nucleolar protein CBFS) (P64').  
 GN CBFS OR YLR175W OR U9470.11.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RC MEDLINE=93330283; Pubmed=8336724;  
 RA Jiang W., Middleton K., Yoon H.-J., Fouquet C., Carbon J.,  
 RA "An essential yeast protein, CBF5p, binds in vitro to centromeres and  
 RT microtubules."  
 RT MOL. CELL. BIOL. 13:4884-4893(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972.  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,

```

RA Favellio A., Fulton L., Gatling S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Riken L., Riles L., Taich A., Trevasakis E., Vignati D.,
RA Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.,
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A
CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH. MAY BE IN
CC SOME WAY ASSOCIATED WITH THE CBF3 110 KDA SUBUNIT (CBF3A).
CC -1- SUBCELLULAR LOCATION: NUCLEUS; NUCLEOLAR.
CC -1- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L12351; AAA34473.1; -
DR EMBL: U17246; AAB67463.1; -
DR PIR: S41853; S41853.
DR SGD: S0004165; CBF3.
DR InterPro: IPR002478; CBF3.
DR InterPro: IPR002501; Trub_N.
DR Pfam: PF01472; PUA; 1.
DR Pfam: PF01509; Trub_N; 1.
DR SMART: SM00359; PUA; 1.
DR Microtubules: Centromere; Repeat; Nuclear protein; DNA-binding.
KM DOMAIN 434 463 10 X 3 AA TANDEM REPEATS OF K-K-[DE].
FT REPEAT 434 436 1.
FT REPEAT 437 439 2.
FT REPEAT 440 442 3.
FT REPEAT 443 445 4.
FT REPEAT 446 448 5.
FT REPEAT 449 451 6.
FT REPEAT 452 454 7.
FT REPEAT 455 457 8.
FT REPEAT 458 460 9.
FT REPEAT 461 463 10.
SQ SEQUENCE 483 AA; 54704 MW; D356B39FDCC32B2D CRC64;

Query Match 67.9%; Score 74; DB 1; Length 483;
Best Local Similarity 70.0%; Pred. No. 0.23;
Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
DB 434 KKEKKKKKKKKKKKKKK 453

RESULT 10
BRD3_HUMAN STANDARD; PRT; 726 AA.
AC 015059; 092645;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Bromodomain-containing protein 3 (RING3-like protein).
GN BRD3 OR RING3L OR KIAA0043.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RX MEDLINE=96051398; Pubmed=7584044;
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,

```

```

RA Seki N., Kawarabayashi Y., Ishikawa K.-I., Tabata S.,
RT "Prediction of the coding sequences of unidentified human genes. II.
RT The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by
RT analysis of cDNA clones from human cell line KG-1."
RL DNA Res. 1:223-229(1994).
RN [2]
RP SEQUENCE OF 363-726 FROM N.A.
RA MEDLINE=96038990; Pubmed=9373153;
RA Thorpe K.L., Gorman P., Thomas C., Sheer D., Trowdale J., Beck S.;
RT "Chromosomal localization, gene structure and transcription pattern of
RT the ORF3 gene, a homologue of the MHC-linked RING3 gene."
RL Gene 200:177-183(1997).
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D26362; BA05393.1; -
DR EMBL: Z81330; CAB03630.1; -
DR MIM: 601541; -
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; bromodomain; 2.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO; 2.
DR PROSITE: PS00633; BROMODOMAIN_1; 2.
DR PROSITE: PS50014; BROMODOMAIN_2; 2.
KM Bromodomain; Repeat; Nuclear protein.
FT DOMAIN 56 115 BROMODOMAIN 1.
FT DOMAIN 326 398 BROMODOMAIN 2.
FT DOMAIN 487 555 LYS-RICH.
FT DOMAIN 676 725 SER-RICH.
FT CONFLICT 465 466 EL -> DY (IN REF. 2).
SQ SEQUENCE 726 AA; 79541 MW; 64F526FC3C1033AA CRC64;

Query Match 67.9%; Score 74; DB 1; Length 726;
Best Local Similarity 70.0%; Pred. No. 0.31;
Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
DB 489 KKKKKKKKKKKKKKKKK 508

RESULT 11
CBF5_CANAL STANDARD; PRT; 479 AA.
AC 043101;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Centromere/microtubule binding protein CBF5 (centromere-binding factor
DE 5) (Nucleolar protein CBF5).
GN CBF5.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OC NCBI_TaxID=5476;
RX MEDLINE=96051398; Pubmed=7584044;
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,

```

CC SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U59149; AAB94297.1; -  
DR InterPro: IPR002478; PUA.  
DR InterPro: IPR002501; TRUB\_N.  
DR Pfam: PF01472; PUA; 1.  
DR Pfam: PF01509; TRUB\_N; 1.  
DR SMART: SM00359; PUA; 1.  
KM Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.  
SO SEQUENCE 479 AA; 54321 MW; 3BAF5104E12C9EB6 CRC64;

Query Match 66.1%; Score 72; DB 1; Length 479;  
Best Local Similarity 70.0%; Pred. No. 0.35;  
Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKK 21  
DB 432 KKKKKKKKKKKKKKKKK 451

RESULT 12  
ID GARP\_PLAFF STANDARD; PRT; 678 AA.  
AC P13616; 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Glutamic acid-rich protein precursor.  
GN GARP.  
OS Plasmodium falciparum (Isolate FC27 / Papua New Guinea).  
CC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
CX NCBI\_TaxID=5837;  
RN [1]  
RP MEDLINE=69040048; PubMed=2903445;  
RA Trigila T., Stahl H.-D., Crewtcher P.E., Silva A., Anders R.F.,  
Kemp D.J.;  
RT "Structure of a Plasmodium falciparum gene that encodes a glutamic  
RT acid-rich protein (GARP)."  
RL Mol. Biochem. Parasitol. 51:199-202(1988).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: J03998; AAA29605.1; -  
DR PIR: A54514; A54514.  
KW Repeat; Malaria; Antigen; Signal.  
FT SIGNAL 1 25  
FT CHAIN 26 678  
FT DOMAIN 120 164  
FT DOMAIN 372 416  
FT DOMAIN 417 441  
FT DOMAIN 576 604  
FT DOMAIN 605 653  
FT DOMAIN 654 663  
FT SEQUENCE 678 AA; 80551 MW; ZA8F85606496EA9E CRC64;

Query Match 66.1%; Score 72; DB 1; Length 678;  
Best Local Similarity 70.0%; Pred. No. 0.46;  
Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKK 21  
DB 132 KKKKKKKKKKKKKKKKK 151

RESULT 13  
ID CG79\_HUMAN STANDARD; PRT; 351 AA.  
AC Q9Y388;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein CG7-79.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20272150; PubMed=10810093;  
RA Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;  
RT "Identification of novel human genes evolutionarily conserved in  
RT Caenorhabditis elegans by comparative proteomics."  
RL Genome Res. 10:703-713(2000).  
CC -1- SIMILARITY: CONTAINS A RNA RECOGNITION MOTIF (RNM).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF151837; AAD34074.1; -  
DR HSP; P09651; 1HA1.  
DR InterPro: IPR000504; RNM.  
DR Pfam: PF00076; RNM; 1.  
DR SMART: SM00360; RNM; 1.  
DR PROSITE: PS0102; RNM; 1.  
DR PROSITE: PS00030; RNM; 1.  
KW Hypothetical protein; RNA-binding.  
FT DOMAIN 36 114  
FT SEQUENCE 351 AA; 39675 MW; 756F862DB192EBE CRC64;

Query Match 65.1%; Score 71; DB 1; Length 351;  
Best Local Similarity 73.7%; Pred. No. 0.35;  
Matches 14; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKK 20  
DB 156 KKKKKKKKKKKKKKKKK 174

RESULT 14  
ID CNG1\_HUMAN STANDARD; PRT; 686 AA.  
AC P29973; Q16485; Q16279;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE cGMP-gated cation channel alpha 1 (CNG channel alpha 1) (CNG1)  
DE (Cyclic nucleotide-gated channel alpha 1) (CNG1)  
DE (Cyclic nucleotide-gated channel alpha 1) (CNG1)  
DE (Rod photoreceptor) (Cyclic nucleotide-gated cation channel 1)  
DE (Rod photoreceptor cGMP-gated channel alpha subunit).  
GN CNGA1 OR CNCG1 OR CNCG.  
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Retina;  
 RX MEDLINE=92210603; PubMed=1372902;  
 RA Piltner S.J., Lee A.K., Altherr M.R., Howard T.A., Seidlin M.F.,  
 RA Hurwitz R.L., Wasmuth J.J., Baehr M.;  
 RT "Primary structure and chromosomal localization of human and mouse  
 RT rod photoreceptor cGMP-gated cation channel.";  
 RL J. Biol. Chem. 267:6257-6262(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Retina;  
 RX MEDLINE=92356211; PubMed=1379636;  
 RA Dhallan R.S., Macke J.P., Eddy R.L., Shows T.B., Reed R.R.,  
 RA Yau K.-W., Nathans J.;  
 RT "Human rod photoreceptor cGMP-gated channel: amino acid sequence,  
 RT gene structure, and functional expression.";  
 RL J. Neurosci. 12:3248-3256(1992).  
 RN [3]  
 RP SEQUENCE OF 313-573 FROM N.A.  
 RX MEDLINE=95175019; PubMed=7532814;  
 RA Distler M., Biel M., Flockert V., Hofmann F.;  
 RT "Expression of cyclic nucleotide-gated cation channels in non-sensory  
 RT tissues and cells.";  
 RL Neuropharmacology 33:1275-1282(1994).  
 RN [4]  
 RP VARIANT ARRP PHE-316, AND VARIANTS GLN-28 AND ASN-114.  
 RX MEDLINE=96036047; PubMed=7479749;  
 RA Dryja T.P., Finn J.T., Peng Y.-W., McGee T.L., Berson E.L., Yau K.-W.;  
 RT "Mutations in the gene encoding the alpha subunit of the rod  
 RT cGMP-gated channel in autosomal recessive retinitis pigmentosa.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:10177-10181(1995).  
 CC -1- FUNCTION: VISUAL SIGNAL TRANSDUCTION IS MEDIATED BY A G-PROTEIN  
 CC COUPLED CASCADE USING CGMP AS SECOND MESSENGER. THIS PROTEIN CAN  
 CC BE ACTIVATED BY CYCLIC GMP WHICH LEADS TO A OPENING OF THE CATION  
 CC CHANNEL AND THEREBY CAUSING A DEPOLARIZATION OF ROD  
 CC PHOTORECEPTORS.  
 CC -1- SUBUNIT: HOMOTETRAMER OR HIGHER OLIGOMER. FORMS HETEROOLIGOMERIC  
 CC COMPLEX WITH CNGB.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- TISSUE SPECIFICITY: ROD CELLS IN THE RETINA.  
 CC -1- DISEASE: DEFECTS IN CNGB1 ARE A CAUSE OF AUTOSOMAL RECESSIVE  
 CC RETINITIS PIGMENTOSA (ARRP). A DISEASE THAT LEADS TO DEGENERATION  
 CC OF RETINAL PHOTORECEPTOR CELLS.  
 CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL  
 CC FAMILY.  
 CC -1- DATABASE: NAME=Mutations of the CNGB1 gene;  
 CC NOTE=Retina International's Scientific Newsletter;  
 CC WWW="http://www.retina-international.com/sci-news/cnbgmut.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 CC -----  
 DR EMBL: M84741; AAA52010.1; ALT\_INIT.  
 DR EMBL: S42457; AAB32778.1; -  
 DR EMBL: S76062; AAD14206.1; -  
 DR PIR: A42161; A42161.  
 DR MIM: 123825; -  
 DR InterPro: IPR000636; Cation\_chan\_non\_lig.  
 DR InterPro: IPR000595; CNMP\_binding.  
 DR Pfam: PF00027; CNMP\_binding.1.  
 DR Pfam: PF00520; Ion\_trans.1.  
 DR SMART: SM00100; CNMP.1.  
 DR PROSITE: PS00888; CNMP\_BINDING\_1; 1.  
 DR PROSITE: PS00889; CNMP\_BINDING\_2; 1.

DR PROSITE: PS0042; CNMP\_BINDING\_3; 1.  
 KW Ionic channel; Ion transport; CAMP-binding; Transmembrane;  
 KW Multigene family; Vision; Disease mutation; Polymorphism;  
 KW Retinitis pigmentosa.  
 FT DOMAIN 1 160  
 FT TRANSSEM 181  
 FT DOMAIN 182 194  
 FT TRANSSEM 195 213  
 FT DOMAIN 214 237  
 FT TRANSSEM 238 257  
 FT DOMAIN 258 295  
 FT TRANSSEM 296 318  
 FT DOMAIN 319 370  
 FT TRANSSEM 371 390  
 FT DOMAIN 391 474  
 FT TRANSSEM 475 495  
 FT DOMAIN 496 686  
 FT NP\_BIND 483 605  
 FT BINDING 542 542  
 FT BINDING 557 557  
 FT CARBOHYD 421 421  
 FT VARIANT 28 28  
 FT VARIANT 114 114  
 FT VARIANT 316 316  
 FT CONFLICT 46 46  
 FT CONFLICT 85 85  
 FT CONFLICT 146 147  
 FT CONFLICT 539 539  
 FT CONFLICT 677 678  
 FT SEQUENCE 686 AA; 79126 MW; E520D216FC97AB6 CRC64;  
 QY 2 KKKKKKKKKKKKKKKKK 21  
 DB 124 KKKKKKKKKKKKKKK 143  
 RESULT 15  
 TCOF\_HUMAN STANDARD; PRT; 1411 AA.  
 AC Q13428; O99408;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Treacher protein (Treacher collins syndrome protein).  
 GN TCOF1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96154183; PubMed=8563749;  
 RA Dixon J., Edwards S.J., Gladwin A.J., Dixon M.J., Loftus S.K.,  
 RA Bonner C.A., Koprivnikar K., Wasmuth J.J.;  
 RT "Positional cloning of a gene involved in the pathogenesis of  
 RT Treacher collins syndrome. The Treacher collins Syndrome  
 RT Collaborative Group.";  
 RL Nat. genet. 12:130-136(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97250498; PubMed=9096354;  
 RA Wise C.A., Chiang L.C., Paznekas W.A., Sharma M., Musy M.M.,  
 RA Ashley J.A., Lovett M., Jabs E.W.;  
 RT "TCOF1 gene encodes a putative nucleolar phosphoprotein that exhibits  
 RT mutations in Treacher Collins syndrome throughout its coding

RT region.";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:3110-3115(1997).  
RN [3]  
RP VARIANTS L-439; V-810; V-1313 AND G-1355, AND VARIANT TCS R-53.  
RX MEDLINE=97195537; PubMed=9042910;  
RA Edwards S.J., Gladwin A.J., Dixon M.J.;  
RT "The mutational spectrum in Treacher Collins syndrome reveals a  
predominance of mutations that create a premature-termination  
codon."  
RL Am. J. Hum. Genet. 60:515-524(1997).  
CC -1- DISEASE: DEFECTS IN TCOF1 ARE THE CAUSE OF TREACHER COLLINS  
SYNDROME (TCS). TCS IS A AUTOSOMAL DOMINANT DISORDER OF  
CRANIOFACIAL DEVELOPMENT THAT OCCURS WITH AN INCIDENCE OF 1/50,000  
LIVE BIRTHS. THE CLINICAL FEATURES OF TCS ARE BILATERALLY  
ASYMMETRICAL AND INCLUDE: (1) ABNORMALITIES OF THE EXTERNAL EARS,  
ATRESIA OF THE EXTERNAL EAR CANALS, AND MALFORMATION OF THE MIDDLE  
EAR OSSICLES, WHICH MAY RESULT IN CONDUCTIVE HEARING LOSS; (2)  
LATERAL DOWARD SLOPING OF PALPEBRAL FISSURES, FREQUENTLY WITH  
COLOBOMAS OF THE LOWER EYELIDS; (3) HYPOPLASIA OF THE MANDIBLE AND  
ZYGOMATIC COMPLEX; (4) CLEFT PALATE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
CC -----  
DR EMBL; U40847; AAC50903.1; -;  
DR EMBL; U76366; AAC51181.1; -;  
DR EMBL; U84664; AAC51185.1; JOINED.  
DR EMBL; U84640; AAC51185.1; JOINED.  
DR EMBL; U84641; AAC51185.1; JOINED.  
DR EMBL; U84642; AAC51185.1; JOINED.  
DR EMBL; U84643; AAC51185.1; JOINED.  
DR EMBL; U84644; AAC51185.1; JOINED.  
DR EMBL; U84645; AAC51185.1; JOINED.  
DR EMBL; U84646; AAC51185.1; JOINED.  
DR EMBL; U84647; AAC51185.1; JOINED.  
DR EMBL; U84648; AAC51185.1; JOINED.  
DR EMBL; U84649; AAC51185.1; JOINED.  
DR EMBL; U84650; AAC51185.1; JOINED.  
DR EMBL; U84651; AAC51185.1; JOINED.  
DR EMBL; U84652; AAC51185.1; JOINED.  
DR EMBL; U84653; AAC51185.1; JOINED.  
DR EMBL; U84654; AAC51185.1; JOINED.  
DR EMBL; U84655; AAC51185.1; JOINED.  
DR EMBL; U84656; AAC51185.1; JOINED.  
DR EMBL; U84657; AAC51185.1; JOINED.  
DR EMBL; U84658; AAC51185.1; JOINED.  
DR EMBL; U84659; AAC51185.1; JOINED.  
DR EMBL; U84660; AAC51185.1; JOINED.  
DR EMBL; U84661; AAC51185.1; JOINED.  
DR EMBL; U84662; AAC51185.1; JOINED.  
DR EMBL; U84663; AAC51185.1; JOINED.  
DR MIM: 154500; -;  
DR InterPro: IPR003993; treacle.  
DR PRINTS: PR01503; TREACLE.  
KW Disease mutation; Polymorphism.  
FT DOMAIN 89 97 POLY-GLU.  
FT DOMAIN 204 207 POLY-SER.  
FT DOMAIN 616 619 POLY-SER.  
FT DOMAIN 919 924 POLY-SER.  
FT DOMAIN 1285 1289 POLY-LYS.  
FT DOMAIN 1375 1386 POLY-LYS.  
FT DOMAIN 1398 1405 POLY-LYS.  
FT VARIANT 53 53 W -> R (IN TCS).  
FT VARIANT 439 439 /FTId=VAR\_005630.  
FT P -> L.  
FT VARIANT 810 810 /FTId=VAR\_005631.  
FT A -> V.  
FT VARIANT 810 810 /FTId=VAR\_005632.

FT VARIANT 1313 1313 A -> V.  
FT FT /FTId=VAR\_005633.  
FT VARIANT 1355 1355 D -> G.  
FT FT /FTId=VAR\_005634.  
SQ SEQUENCE 1411 AA; 144312 MW; 3880203D985C2699 CRC64;  
  
Query Match 64.7%; Score 70.5; DB 1; Length 1411;  
Best Local Similarity 54.8%; Pred. No. 1.1;  
Matches 17; Conservative 2; Mismatches 1; Indels 11; Gaps 1;  
  
QY 2 KKKKKKKKKK-----KKKKKK 21  
DB 1375 KKKKKKKKKAKASTKDESPSKKKKKK 1405  
  
Search completed: July 1, 2002, 06:31:33  
Job time: 670 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:19:28 ; Search time 83.98 Seconds

(without alignments)  
43.259 Million cell updates/sec

Title: US-09-461-684-1

Perfect score: 109

Sequence: 1 CXXXXXXXXXXXXXXXXXXXX 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_19:\*

1: sp.archaea:\*

2: sp.bacteria:\*

3: sp.fungi:\*

4: sp.human:\*

5: sp.invertebrate:\*

6: sp.mammal:\*

7: sp.mbc:\*

8: sp.organelle:\*

9: sp.phage:\*

10: sp.plant:\*

11: sp.protent:\*

12: sp.virus:\*

13: sp.vertebrate:\*

14: sp.unclassified:\*

15: sp.virus:\*

16: sp.bacteriophage:\*

17: sp.archaeo:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	91.7	129	3 Q9P529	Q9P529 neurospora
2	100	91.7	129	11 Q35807	Q35807 ratulus norv
3	100	91.7	168	4 Q9H5V5	Q9H5V5 homo sapien
4	100	91.7	215	11 Q64075	Q64075 rattus norv
5	100	91.7	260	10 Q9L629	Q9L629 arabidopsis
6	100	91.7	380	4 Q9NT34	Q9NT34 homo sapien
7	100	91.7	517	10 Q9LXR2	Q9LXR2 arabidopsis
8	100	91.7	531	6 Q9SLV5	Q9SLV5 macaca fasc
9	100	91.7	720	4 Q9H607	Q9H607 homo sapien
10	97	85.0	667	4 Q9HC48	Q9HC48 homo sapien
11	90	82.6	257	4 Q9H5Y3	Q9H5Y3 homo sapien
12	90	82.6	483	5 Q96148	Q96148 plasmidium
13	90	82.6	686	4 Q9NKF0	Q9NKF0 homo sapien
14	89	81.7	83	11 Q9ER82	Q9ER82 mus musculu
15	87	79.8	107	10 Q9LOR6	Q9LOR6 arabidopsis
16	81	74.3	4550	5 Q77336	Q77336 plasmidium

17	80	73.4	130	4 Q9H7E1	Q9H7E1 homo sapien
18	80	73.4	213	11 P97762	P97762 mus musculu
19	80	73.4	784	5 Q77339	Q77339 plasmidium
20	79	72.5	260	11 Q991X5	Q991X5 mus musculu
21	78	71.6	263	4 Q9BU76	Q9BU76 homo sapien
22	78	71.6	407	4 Q00536	Q00536 homo sapien
23	78	71.6	529	4 Q9NPX4	Q9NPX4 homo sapien
24	78	71.6	560	10 Q65334	Q65334 plasmidium
25	78	71.6	628	4 Q9H5M5	Q9H5M5 homo sapien
26	78	71.6	1591	11 P97868	P97868 mus musculu
27	78	71.6	1616	4 Q96PH3	Q96PH3 homo sapien
28	77.5	71.1	353	4 Q9HBQ9	Q9HBQ9 homo sapien
29	77	70.6	103	5 Q9N8X7	Q9N8X7 trypanosoma
30	77	70.6	229	11 Q9JMG3	Q9JMG3 mus musculu
31	77	70.6	266	10 Q9X134	Q9X134 arabidopsis
32	77	70.6	402	4 Q9H4Q7	Q9H4Q7 homo sapien
33	77	70.6	577	3 Q9C1W8	Q9C1W8 schizosach
34	77	70.6	749	5 Q967D9	Q967D9 drosophila
35	77	70.6	3724	5 Q77320	Q77320 plasmidium
36	76	69.7	581	5 Q96549	Q96549 dictyosteli
37	76	69.7	812	5 Q21827	Q21827 caenorhabdi
38	76	69.7	1629	5 Q9U0K9	Q9U0K9 plasmidium
39	75	68.8	2086	11 Q9QXP5	Q9QXP5 mus musculu
40	75	68.8	2404	11 Q9QX47	Q9QX47 mus musculu
41	74	67.9	191	5 Q45682	Q45682 caenorhabdi
42	74	67.9	198	16 Q97E51	Q97E51 clostridium
43	74	67.9	244	11 Q9CWL7	Q9CWL7 mus musculu
44	74	67.9	256	11 Q9CUI4	Q9CUI4 mus musculu
45	74	67.9	550	5 Q9VVF1	Q9VVF1 drosophila

## ALIGNMENTS

RESULT 1

Q9P529 ID Q9P529 PRELIMINARY; PRT; 128 AA.

AC Q9P529;

DT 01-Oct-2000 (TReMBLrel. 15, Created)

DT 01-Dec-2001 (TReMBLrel. 19, Last sequence update)

DT 01-Dec-2001 (TReMBLrel. 19, Last annotation update)

DE HYPOTHETICAL 15.2 KDA PROTEIN.

GN B24H17.160.

CN Neurospora crassa.

OS Neurospora crassa.

CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

CC Sordariales; Sordariaceae; Neurospora.

OX NCBI\_Taxid=5141;

RN [1]

RP SEQUENCE FROM N.A.

RA Schulte U., Alpn V., Hohnsels J., Brandt P., Fartmann B., Holland R.,

RA Nyakatura G., Mewes H.W., Mannhaupt G.,

RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA German Neurospora genome project;

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL356815; CAB92638.2; -

KK Hypothetical protein

SO SEQUENCE 128 AA; 15157 MW; 8C7C65C3DFE70765 CRC64;

Query Match 91.7%; Score 100; DB 3; Length 128;  
Best Local Similarity 100.0%; Pred. No. 9,6e-05;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KXXXXXXXXXXXXXXXXXXXX 21  
|XXXXXXXXXXXXXXXXXXXX|

Db 71 KXXXXXXXXXXXXXXXXXXXX 90  
|XXXXXXXXXXXXXXXXXXXX|

RESULT 2  
Q35807 ID Q35807 PRELIMINARY; PRT; 129 AA.

AC 035807;  
 DT 01-JAN-1998 (TReMBLrel. 05, Created)  
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE MICROVASCULAR ENDOTHELIAL DIFFERENTIATION PROTEIN 2.  
 GN MDC2.  
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NC NCBL\_TaxID=10116;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-EPIDIDYMIS;  
 RA MEDLINE=98172708; PubMed=9511718;  
 RT "Differential expression of osteopontin, PC4, and CEC5, a novel mRNA species, during in vitro angiogenesis.";  
 RT Exp. Cell Res. 239:1-10(1998).  
 DR EMBL; Y08769; CAA70022.1; -;  
 DR InterPro; IPR000719; Euk.pkinase.  
 DR Pfam; PF00069; pkinase.1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW ATP-binding; Transferase.  
 SQ SEQUENCE 129 AA; 15080 MW; 38102272B8E2EDB4 CRC64;

Query Match 91.7%; Score 100; DB 11; Length 129;  
 Best Local Similarity 100.0%; Pred. No. 9.7e-05;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21  
 DB 85 KKKKKKKKKKKKKKKKKKK 104

RESULT 3  
 ID Q9H5V6 PRELIMINARY; PRT; 168 AA.  
 DT 01-MAR-2001 (TReMBLrel. 16, Created)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
 DE CDNA: FLJ22976 FIS, CLONE KAT11222 (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBL\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Matnabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,  
 RA Suzuki Y., Ohtsuka M., Nishi T., Shibahara T., Tanaka T.,  
 RA Nakamura Y., Isogai T., Sugano S.;  
 RT "NEDD human cDNA sequencing project."  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK026629; BAB15513.1; -;  
 FT NON\_TER 168  
 SQ SEQUENCE 168 AA; 19549 MW; A19DBD195F8A1A90 CRC64;

Query Match 91.7%; Score 100; DB 4; Length 168;  
 Best Local Similarity 100.0%; Pred. No. 0.00012;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21  
 DB 140 KKKKKKKKKKKKKKKKKKK 159

RESULT 4  
 ID Q64075 PRELIMINARY; PRT; 215 AA.  
 AC Q64075;  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE NCLEOPORIN P62 HOMOLOG PROTEIN (FRAGMENT).  
 OS Rattus sp.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBL\_TaxID=10118;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=95151924; PubMed=7849178;  
 RA Wang Z.Q., Akmal K.M., Kim K.H.;  
 RT "An unusual nucleoporin-related messenger ribonucleic acid is present in the germ cells of rat testis."  
 RT Biol. Reprod. 51:1022-1030(1994).  
 DR EMBL; S75997; AAB3384.1; -;  
 KW Porin.  
 FT NON\_TER 1  
 SQ SEQUENCE 215 AA; 24593 MW; 098251C97A8FBD8 CRC64;

Query Match 91.7%; Score 100; DB 11; Length 215;  
 Best Local Similarity 100.0%; Pred. No. 0.00014;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21  
 DB 35 KKKKKKKKKKKKKKKKKKK 54

RESULT 5  
 ID Q9LG29 PRELIMINARY; PRT; 260 AA.  
 AC Q9LG29;  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE GENOMIC DNA, CHROMOSOME 3, BAC CLONE:FLD9.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBL\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-COLUMBIA;  
 RA Nakamura Y.;  
 RT "Structural Analysis of Arabidopsis thaliana Chromosome 3. III."  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP002460; BA97098.1; -;  
 DR InterPro; IPR001386; Linker\_histone.  
 DR PRINTS; PR00624; HISTONEH9.  
 SQ SEQUENCE 260 AA; 33307 MW; 43E2394CB8131143 CRC64;

Query Match 91.7%; Score 100; DB 10; Length 260;  
 Best Local Similarity 100.0%; Pred. No. 0.00016;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21  
 DB 7 KKKKKKKKKKKKKKKKKKK 26

RESULT 6  
 ID Q9NT34 PRELIMINARY; PRT; 380 AA.  
 AC Q9NT34;  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)  
 DE HYPOTHETICAL 42.7 KDA PROTEIN (FRAGMENT).  
 GN DKF2P434I1120.  
 OS Homo sapiens (Human).



```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Ohtenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL, AL137556; CAB70810.1; -
KM Hypothetical protein.
FT NON_TER 380
SQ SEQUENCE 380 AA; 42689 MW; 67F5DD101346AFB CRC64;

Query Match
Best Local Similarity 100.0%; Score 100; DB 4; Length 380;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
DB 355 KKKKKKKKKKKKKKKKKKK 374

RESULT 7
Q9LXR2 PRELIMINARY; PRT; 517 AA.
AC Q9LXR2;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE HYPOTHETICAL 59.7 KDA PROTEIN.
GN T20N10_250.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,
RA Rued S., Lemcke K., Mayer K.F.X., Quettier F., Salanoubat M.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL, AL135303; CAB88307.1; -
DR InterPro: IPR001810; F-box.
DR Pfam: PF00646; F-box; 1.
DR SMART: SM00256; FBOX; 1.
DR PROSITE: PS50181; FBOX; 1.
KM Hypothetical protein.
SQ SEQUENCE 517 AA; 59689 MW; EC6D957D01F86E70 CRC64;

Query Match
Best Local Similarity 91.7%; Score 100; DB 10; Length 517;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
DB 444 KKKKKKKKKKKKKKKKKKK 463

RESULT 8
Q95LV6 PRELIMINARY; PRT; 531 AA.
AC Q95LV6;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 61.4 KDA PROTEIN (FRAGMENT).
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA Terao K., Sugano S.;
RT "Isolation of novel full-length cDNA clones from macaque testis cDNA
RT libraries."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL, AB071085; BAB64479.1; -
KM Hypothetical protein.
FT NON_TER 531
SQ SEQUENCE 531 AA; 61389 MW; B55996B4F5CDD60C CRC64;

Query Match
Best Local Similarity 100.0%; Score 100; DB 6; Length 531;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
DB 502 KKKKKKKKKKKKKKKKKKK 521

RESULT 9
Q9H607 PRELIMINARY; PRT; 720 AA.
AC Q9H607;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CNA: FLJ21979 FIS. CLONE HEP06065 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL, AK025632; BAB15196.1; -
FT NON_TER 720
SQ SEQUENCE 720 AA; 84029 MW; A86586FEAA953D0B CRC64;

Query Match
Best Local Similarity 91.7%; Score 100; DB 4; Length 720;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
DB 692 KKKKKKKKKKKKKKKKKKK 711

RESULT 10
Q9HC48 PRELIMINARY; PRT; 667 AA.
AC Q9HC48;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CTCL TUMOR ANTIGEN SE2-5 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
```

RP SEQUENCE FROM N.A.  
 RC TISSUE=TESTIS;  
 RX MEDLINE=21143360; PubMed=1149944;  
 RA Eichmuller S., Usener D., Dummer R., Stein A., Thiel D.,  
 R Schandendorf D.;  
 RT "Serological detection of cutaneous T-cell lymphoma-associated  
 antigens";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:629-634(2001).  
 DR EMBL: AF177228; FAG3676.1; -.  
 DR HSSP: Q12923; 3PDZ.  
 DR InterPro: IPR001478; PDZ.  
 DR Pfam: PF00595; PDZ; 2.  
 DR SMART: SM00228; PDZ; 2.  
 DR PROSITE: PSS0106; PDZ; 2.  
 DR NON\_TER 1  
 FT NON\_TER 667  
 SQ SEQUENCE 667 AA; 73499 MW; C653EC16802BAE02 CRC64;

Query Match 89.0%; Score 97; DB 4; Length 667;  
 Best Local Similarity 95.0%; Pred. No. 0.00063;  
 Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21  
 DB 648 RKKKKKKKKKKKKKKKKKK 667

RESULT 11  
 Q9H5Y3 PRELIMINARY; PRT; 257 AA.  
 AC Q9H5Y3;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE CDNA: FLJ22789 FIS, CLONE KAI1A171 (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TISSUE=TESTIS;  
 RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,  
 Nakajima Y., Mizuno T., Morinaga M., Taniguchi A., Fujiwara T., Ono T.,  
 Yamada K., Fujii Y., Ozaki K., Hiroo M., Ohmori Y., Ota T., Suzuki Y.,  
 RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,  
 RA Isogai T., Sugano S.;  
 RT "NDO human cDNA sequencing project";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK026442; BAB15484.1; -.  
 DR NON\_TER 257  
 FT NON\_TER 257  
 SQ SEQUENCE 257 AA; 28778 MW; 100DEA6ABA521868 CRC64;

Query Match 82.6%; Score 90; DB 4; Length 257;  
 Best Local Similarity 100.0%; Pred. No. 0.0019;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 19  
 DB 240 KKKKKKKKKKKKKKKKKKK 257

RESULT 12  
 O96148 PRELIMINARY; PRT; 483 AA.  
 AC O96148;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DE 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)  
 DE HYPOTHEITICAL 57.8 KDA PROTEIN.  
 GN PRB0235W.

OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99021743; PubMed=9804551;  
 RA Gardner M.J., Jetelelin H., Carucci D.J., Cummings L.M., Aravind L.,  
 RA Koonin E.V., Shalimov S., Mason T., Yu K., Fujii C., Pederson J.,  
 RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Perera M.,  
 RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,  
 RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;  
 RT "Chromosome 2 sequence of the human malaria parasite Plasmodium  
 falciparum";  
 RL Science 282:1126-1132(1999).  
 DR EMBL: AE001362; AAC71836.1; -.  
 DR KW Hypothetical protein.  
 SQ SEQUENCE 483 AA; 57785 MW; 9332C1EC9A4DECFB CRC64;

Query Match 82.6%; Score 90; DB 5; Length 483;  
 Best Local Similarity 85.0%; Pred. No. 0.0029;  
 Matches 17; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21  
 DB 449 KKKKKKKKKKKKKKKKKKK 468

RESULT 13  
 Q9NXP0 PRELIMINARY; PRT; 666 AA.  
 AC Q9NXP0;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE CDNA: FLJ20288 FIS, CLONE HEP04414 (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,  
 Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,  
 RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
 RT "NDO human cDNA sequencing project";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK000295; BAA91063.1; -.  
 DR HSSP: O00420; IAWC.  
 DR InterPro: IPR002110; ANK.  
 DR Pfam: PF00023; ANK; 14.  
 DR SMART: SM00248; ANK; 14.  
 DR PROSITE: PSS0088; ANK\_REPEAT; 13.  
 DR PROSITE: PSS0297; ANK\_REPEAT\_REGION; 1.  
 DR KW ANK repeat; Repeat.  
 DR NON\_TER 686  
 FT NON\_TER 686  
 SQ SEQUENCE 686 AA; 73615 MW; AE139C0212E0FAD22 CRC64;

Query Match 82.6%; Score 90; DB 4; Length 686;  
 Best Local Similarity 85.0%; Pred. No. 0.0037;  
 Matches 17; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21  
 DB 666 KKKKKKKKKKKKKKKKKKK 685

RESULT 14  
 Q9ER82 PRELIMINARY; PRT; 83 AA.  
 AC Q9ER82;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)

Search completed: July 1, 2002, 06:30:54  
Job time: 686 sec

DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DE 01-MAR-2001 (TREMBLrel. 16, last annotation update)  
DE SR528 PROTEIN (FRAGMENT).  
GN SR528.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/10; TISSUE=SKLETAL MUSCLE;  
RA Kemp T.J., Sadusky T.J., Carey N., Coulton G.R.;  
RT "Murine genes identified from skeletal muscle mRNA which had undergone  
RT 7 days of passive stretch in vivo."  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ250693; CAC08505.1; -  
FT NON\_TER  
SQ SEQUENCE 83 AA; 10093 MW; 308BB2113D02DE92 CRC64;

Query Match 81.7%; Score 89; DB 11; Length 83;  
Best Local Similarity 85.0%; Pred. No. 0.0011;  
Matches 17; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21  
DB 49 KKKKKKKKKKKKKKKKKKK 68

RESULT 15  
ID O9LOF6 PRELIMINARY; PRT; 107 AA.  
AC O9LOF6;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)  
DE F1504.29.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Cheuk R., Shin P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
RA Khan S., Kim C., Altafi H., Bel B., Chin C., Chlou J., Choi E., Choi T.,  
RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,  
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,  
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC007887; AAF79354.1; -  
SQ SEQUENCE 107 AA; 13427 MW; 91091726AED96615 CRC64;

Query Match 79.8%; Score 87; DB 10; Length 107;  
Best Local Similarity 94.4%; Pred. No. 0.0022;  
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 19  
DB 29 KKKKKKKKKKKKKKKKKKK 46









CC disease). The polynucleotide sequences of the invention may also be  
CC used in gene therapy. AA018154-AA018281 represent novel DNA-binding  
CC proteins.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 58 AA;  
  
Query Match 100.0%; Score 109; DB 22; Length 58;  
Best Local Similarity 100.0%; Pred. No. 5e-06; 0;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CKNKKKKKKKKKKKKKKKKKK 21  
|||||  
DB 35 CKKKKKKKKKKKKKKKKKKK 55  
  
RESULT 4  
AA011210  
ID AA011210 standard; Protein: 70 AA.  
XX  
XX AA011210;  
XX  
XX 06-NOV-2001 (first entry)  
XX  
XX Human polypeptide SEQ ID NO 25102.  
XX  
XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;  
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
XX tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX nervous system disorders; arthritis; inflammation.  
XX  
XX Homo sapiens.  
XX  
XX WO200164835-A2.  
XX  
XX 07-SEP-2001.  
XX  
XX 26-FEB-2001; 2001WO-US04927.  
XX  
XX 28-FEB-2000; 2000US-0515126.  
XX 18-MAY-2000; 2000US-0577409.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI; 2001-514838/56.  
XX N-PSDB; AA191141.  
XX  
XX Isolated nucleic acids and polypeptides, useful for preventing  
XX diagnosing and treating e.g. leukaemia, inflammation and immune  
XX disorders -  
XX  
XX Claim 20; SEQ ID NO 25102; 1399pp + Sequence Listing; English.  
XX  
XX The invention relates to human polynucleotides (AA179941-AA193841) and  
XX the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
XX cytokine, cell proliferation or cell differentiation or which may induce  
XX production of other cytokines in other cell populations. The  
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or  
XX peptide therapy. The polypeptides have various cytokine-like activities,  
XX e.g. stem cell growth factor activity, haematopoiesis regulating  
XX activity, tissue growth factor activity, immunomodulatory activity and  
XX activin/inhibin activity and may be useful in the diagnosis and/or  
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and  
XX inflammation.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 70 AA;  
  
Query Match 100.0%; Score 109; DB 22; Length 70;  
Best Local Similarity 100.0%; Pred. No. 5.8e-06;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CKNKKKKKKKKKKKKKKKKKK 21  
|||||  
DB 30 CKKKKKKKKKKKKKKKKKKK 50  
  
RESULT 5  
AA003766  
ID AA003766 standard; Protein: 81 AA.  
XX  
XX AA003766;  
XX  
XX 06-NOV-2001 (first entry)  
XX  
XX Human polypeptide SEQ ID NO 17658.  
XX  
XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;  
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
XX tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX nervous system disorders; arthritis; inflammation.  
XX  
XX Homo sapiens.  
XX  
XX WO200164835-A2.  
XX  
XX 07-SEP-2001.  
XX  
XX 26-FEB-2001; 2001WO-US04927.  
XX  
XX 28-FEB-2000; 2000US-0515126.  
XX 18-MAY-2000; 2000US-0577409.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI; 2001-514838/56.  
XX N-PSDB; AA183697.  
XX  
XX Isolated nucleic acids and polypeptides, useful for preventing  
XX diagnosing and treating e.g. leukaemia, inflammation and immune  
XX disorders -  
XX  
XX Claim 20; SEQ ID NO 17658; 1399pp + Sequence Listing; English.  
XX  
XX The invention relates to human polynucleotides (AA179941-AA193841) and  
XX the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
XX cytokine, cell proliferation or cell differentiation or which may induce  
XX production of other cytokines in other cell populations. The  
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or  
XX peptide therapy. The polypeptides have various cytokine-like activities,  
XX e.g. stem cell growth factor activity, haematopoiesis regulating  
XX activity, tissue growth factor activity, immunomodulatory activity and  
XX activin/inhibin activity and may be useful in the diagnosis and/or  
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and  
XX inflammation.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 81 AA;  
  
Query Match 100.0%; Score 109; DB 22; Length 81;  
Best Local Similarity 100.0%; Pred. No. 5.4e-06;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



OY 1 CXXXXXXXXXXXXXXXXX 21  
DB 48 CXXXXXXXXXXXXXXXXX 68

RESULT 6  
AA011849  
ID AA011849 standard; Protein; 36 AA.  
XX  
AC AA011849;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 25741.  
XX

KW Human: cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

PN WO200164835-A2.

PD 07-SEP-2001.

PF 26-FEB-2001; 2001WO-US04927.

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

XX MPI; 2001-514838/56.

DR N-PDB; AA191780.

PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -

PS Claim 20; SEQ ID NO 25741; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AA179941-AA193841) and  
XX the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
XX cytokine, cell proliferation or cell differentiation or which may induce  
XX production of other cytokines in other cell populations. The  
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or  
XX peptide therapy. The polypeptides have various cytokine-like activities,  
XX e.g. stem cell growth factor activity, haematopoiesis regulating  
XX activity, tissue growth factor activity, immunomodulatory activity and  
XX activin/inhibin activity and may be useful in the diagnosis and/or  
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and  
XX inflammation.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 36 AA;

Query Match 96.3%; Score 105; DB 22; Length 36;  
Best Local Similarity 95.2%; Pred. No. 9.6e-06;  
Matches 20; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 CXXXXXXXXXXXXXXXXX 21  
DB 11 CXXXXXXXXXXXXXXXXX 31

RESULT 7  
AA012447

ID AA012447 standard; Protein; 47 AA.

XX  
AC AA012447;

DT 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 26339.

KW Human: cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

PN WO200164835-A2.

PD 07-SEP-2001.

PF 26-FEB-2001; 2001WO-US04927.

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

XX MPI; 2001-514838/56.

DR N-PDB; AA192378.

PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -

PS Claim 20; SEQ ID NO 26339; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AA179941-AA193841) and  
XX the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
XX cytokine, cell proliferation or cell differentiation or which may induce  
XX production of other cytokines in other cell populations. The  
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or  
XX peptide therapy. The polypeptides have various cytokine-like activities,  
XX e.g. stem cell growth factor activity, haematopoiesis regulating  
XX activity, tissue growth factor activity, immunomodulatory activity and  
XX activin/inhibin activity and may be useful in the diagnosis and/or  
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and  
XX inflammation.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 47 AA;

Query Match 96.3%; Score 105; DB 22; Length 47;  
Best Local Similarity 95.2%; Pred. No. 1.2e-05;  
Matches 20; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 CXXXXXXXXXXXXXXXXX 21  
DB 12 CXXXXXXXXXXXXXXXXX 32

RESULT 8

ID AA018239 standard; Protein; 66 AA.

XX AA018239;

DT 21-NOV-2001 (first entry)

DE Novel human DNA-binding protein #86.







PT disorders -  
XX  
PS Claim 20; SEQ ID NO 25106; 1399pp + Sequence Listing; English.  
XX  
CC The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 68 AA;  
  
Query Match 94.5%; Score 103; DB 22; Length 68;  
Best Local Similarity 95.2%; Pred. No. 2.6e-05;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 CXXXXXXXXXXXXXXXXX 21  
|XXXXXXXXXXXXXXXXX  
Db 17 cXXXXXXXXXXXXXXXXX 37  
  
RESULT 13  
AA000291  
ID AA000291 standard; Protein; 124 AA.  
XX  
AC AA000291;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 14183.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PE 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
XX  
PR 18-MAY-2000; 2000US-0577409.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
XX  
DR WPI: 2001-514838/56.  
XX  
DR N-PSDB: AA180222.  
XX  
PT Isolated nucleic acids and polypeptides, useful for preventing  
XX diagnosing and treating e.g. leukaemia, inflammation and immune  
XX disorders -  
XX  
PS Claim 20; SEQ ID NO 14183; 1399pp + Sequence Listing; English.  
XX  
CC The invention relates to human polynucleotides (AA179941-AA193841) and  
XX the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
XX cytokine, cell proliferation or cell differentiation or which may induce

CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 124 AA;  
  
Query Match 94.5%; Score 103; DB 22; Length 124;  
Best Local Similarity 95.2%; Pred. No. 4.2e-05;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 CXXXXXXXXXXXXXXXXX 21  
|XXXXXXXXXXXXXXXXX  
Db 20 cXXXXXXXXXXXXXXXXX 40  
  
RESULT 14  
AA012203  
ID AA012203 standard; Protein; 41 AA.  
XX  
AC AA012203;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 26095.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PE 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
XX  
PR 18-MAY-2000; 2000US-0577409.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
XX  
DR WPI: 2001-514838/56.  
XX  
DR N-PSDB: AA192134.  
XX  
PT Isolated nucleic acids and polypeptides, useful for preventing  
XX diagnosing and treating e.g. leukaemia, inflammation and immune  
XX disorders -  
XX  
PS Claim 20; SEQ ID NO 26095; 1399pp + Sequence Listing; English.  
XX  
CC The invention relates to human polynucleotides (AA179941-AA193841) and  
XX the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
XX cytokine, cell proliferation or cell differentiation or which may induce  
XX production of other cytokines in other cell populations. The  
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or  
XX peptide therapy. The polypeptides have various cytokine-like activities,  
XX e.g. stem cell growth factor activity, haematopoiesis regulating  
XX activity, tissue growth factor activity, immunomodulatory activity and  
XX activin/inhibin activity and may be useful in the diagnosis and/or  
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC Inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

Sequence 41 AA;

Query Match 93.6%; Score 102; DB 22; Length 41;  
Best Local Similarity 95.2%; Pred. No. 2.3e-05;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CXXXXXXXXXXXXXXXXXXXX 21  
|XXXXXXXXXXXXXXXXXXXX|  
Db 4 CXXXXXXXXXXXXXXXXXXXX 24

## RESULT 15

AA012187  
ID AA012187 standard; Protein; 60 AA.

AC AA012187;

DT 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 26079.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; Leukaemia;  
KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

PN WO200164835-A2.

PD 07-SEP-2001.

PE 26-FEB-2001; 2001WO-US04927.

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Lin C, Drmanac RT;

DR WPI; 2001-514838/56.

DR N-PSDB; AA192118.

PT Isolated nucleic acids and polypeptides, useful for preventing  
PT disorders and treating e.g. leukaemia, inflammation and immune  
PT disorders -

PS Claim 20; SEQ ID NO 26079; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 60 AA;

Query Match 93.6%; Score 102; DB 22; Length 60;  
Best Local Similarity 95.2%; Pred. No. 3.1e-05;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CXXXXXXXXXXXXXXXXXXXX 21  
|XXXXXXXXXXXXXXXXXXXX|  
Db 26 CXXXXXXXXXXXXXXXXXXXX 46

Search completed: July 1, 2002, 06:19:22  
Job time: 509 sec

Mon Jul 1, 13:54:41 2002

us-09-461-684-1.rag

---

Page 11





OY 2 EAAAAAEEAAAAEEAAAAA 25  
 DB 115 QAEDAKAAAEAEKAAAEAAAAA 138

## RESULT 3

cgcr-1 protein - Chlamydomonas reinhardtii (fragment)

C:Species: Chlamydomonas reinhardtii

C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 03-Nov-2000

C:Accession: S19114

R:Makarchuk, W.W.; Mueller, F.W.; Beck, C.F.

Plant Mol. Biol. 18, 143-146, 1992

A:Title: Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex arrangements

A:Reference number: S19113; MUID:92119224

A:Accession: S19114

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-205 <WAK>

A:Cross-references: EMBL:X17207

C:Superfamily: phage lambda hypothetical protein 401

Query Match 65.1%; Score 71; DB 2; Length 205;  
 Best Local Similarity 75.0%; Pred. No. 0.91;  
 Matches 18; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAAEEAAAAEEAAAAA 25  
 DB 49 EAAAAAQAEEAAAAEEAAAAA 72

## RESULT 4

gene C98 protein - rape

C:Species: Brassica napus (rape)

C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Sep-1997

C:Accession: S24960

R:Roberts, M.R.; Hodge, R.; Sorensen, A.; Ross, J.; Murphy, D.J.; Draper, R.; Scott, R.

submitted to the EMBL data library, July 1992

A:Description: A new class of Brassica napus oleosin genes specific to the male gametoph

A:Reference number: S24960

A:Accession: S24960

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-183 <ROB>

A:Cross-references: EMBL:X67142; NID:917792; PID:917793

Query Match 64.2%; Score 70; DB 2; Length 183;  
 Best Local Similarity 78.3%; Pred. No. 1;  
 Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 AAAAAEAAAAEEAAAAEEAAAAA 25  
 DB 151 AAPAAEPAPAAEAPAAEAPAA 173

## RESULT 5

engrailed homeodomain-containing protein En-1 - mouse

N:Alternate names: homeotic protein En-1

C:Species: Mus musculus (house mouse)

C:Date: 01-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 24-Sep-1999

C:Accession: A48423; S13009; A26629; A24778

R:Logan, C.; Hanks, M.C.; Noble-Topham, S.; Nallanathan, D.; Provart, N.J.; Joyner, A.I.

Dev. Genet. 13, 345-358, 1992

A:Title: Cloning and sequence comparison of the mouse, human, and chicken engrailed gene

A:Reference number: A48423; MUID:93183339

A:Accession: A48423

A:Status: preliminary; not compared with conceptual translation

A:Residues: 1-401 <LOC>

A:Experimental source: CD-1, embryo

A:Note: Sequence extracted from NCBI backbone (NCBI:126620)

R:Holland, P.W.H.; Williams, N.A.

FEBS Lett. 277, 250-252, 1990

A:Title: Conservation of engrailed-like homeobox sequences during vertebrate evolution

A:Reference number: S13009; MUID:91099509

A:Accession: S13009

A:Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 321-380 <HOL>

R:Joyner, A.L.; Martin, G.R.

Genes Dev. 1, 29-38, 1987

A:Title: En-1 and En-2, two mouse genes with sequence homology to the Drosophila engr

A:Reference number: A31620; MUID:88112776

A:Accession: A26629

A:Molecule type: DNA; mRNA

A:Residues: 278-401 <JOY>

A:Cross-references: GB:Y00201; GB:M11987; NID:549587; PIDN:CAA68361.1; PID:9669105

R:Joyner, A.L.; Kornberg, T.; Coleman, K.G.; Cox, D.R.; Martin, G.R.

Cell 43, 29-37, 1985

A:Title: Expression during embryogenesis of a mouse gene with sequence homology to t

A:Reference number: A24778; MUID:86079501

A:Accession: A24778

A:Molecule type: DNA

A:Residues: 311-401 <JO2>

C:Genetics:

A:Gene: en-1

A:Map position: 1

C:Superfamily: unassigned homeobox proteins; homeobox homology

C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

F:313-369/Domain: homeobox homology <HOX>

Query Match 64.2%; Score 70; DB 2; Length 401;  
 Best Local Similarity 78.3%; Pred. No. 1.9;  
 Matches 18; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 3 AAAAAEAAAAEEAAAAEEAAAAA 25  
 DB 207 AAAAAAATAAAAVAAAAAAS 229

## RESULT 6

S02376

antifreeze protein precursor - yellowtail flounder

C:Species: Limanda ferruginea (yellowtail flounder)

C:Date: 01-Dec-1989 #sequence\_revision 01-Dec-1989 #text\_change 24-Oct-2000

C:Accession: S02376

R:Scott, G.K.; Davies, P.L.; Shears, M.A.; Fletcher, G.L.

Eur. J. Biochem. 168, 629-633, 1987

A:Title: Structural variations in the alanine-rich antifreeze proteins of the Pleurc

A:Reference number: S02376; MUID:88029483

A:Accession: S02376

A:Molecule type: mRNA

A:Residues: 1-97 <SCO>

A:Cross-references: EMBL:X06356; NID:964041; PIDN:CAA29655.1; PID:964042

A:Note: part of this sequence, including the amino end of the mature protein, was co

C:Superfamily: antifreeze protein

C:Keywords: antifreeze

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-48/Domain: propeptide #status predicted <PRO>

F:49-96/Product: antifreeze protein #status predicted <MAT>

Query Match 63.3%; Score 69; DB 2; Length 97;  
 Best Local Similarity 70.8%; Pred. No. 0.82;  
 Matches 17; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAAEEAAAAEEAAAAA 25  
 DB 53 DAAAAAAATAAAKAAADTAA 76



QY	2	EAAAAAEAAAAEAAAAEAAAAA	25
		:   : : : : : : :	
Db	366	QAAQAQAQAQAQAQAQAQAQAQA	389

```

RESULT 12
533225
ecdysone-induced Protein E74A - fruit fly (Drosophila virilis)
C:Species: Drosophila virilis
C:Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 21-Feb-1997
C:Accession: B53225
R:Jones, C.W.; Dalton, M.W.; Townley, L.H.
Genetics 127, 535-543, 1991
A:Title: Interspecific comparisons of the structure and regulation of the Drosophila ecd
A:Reference number: A53225; MUID:91200627
A:Accession: B53225
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-873 <GR>
A:Cross-references: GB:X59493
C:Genetics:
A:Gene: FlyBase:Dftr/Efp74EF
A:Cross-references: FlyBase:FBgn0013076
C:Superfamily: ets DNA-binding domain homology
:779-859/domain: ets DNA-binding domain homology <ETS>

```

```

RESULT 13
T13154
Polycomb protein enhancer - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13154
R:Stankunas, K.; Berger, J.; Ruse, C.; Sinclair, D.A.; Randazzo, F.; Brock, H.W.
Development 125, 4055-4066, 1998
A:Title: The enhancer of polycomb gene of Drosophila encodes a chromatin protein
conserv
A:Reference number: 217611; MUID:98407961
A:Accession: T13154
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2023 <STA>
A:Cross-references: EMBL:AF079764; NID:g3757889; PID:g3757890; PIDN:AAC64271.1
A:Experimental source: Imaginal disc
C:Genetics:
A:Gene: E(Pc)
A:Cross-references: flybase:FBgn0000581
A:Map position: 2

```

RESULT 14  
JC5273  
paired type homeobox protein, NBP - human  
C:Species: Homo sapiens (man)  
C:Date: 16-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 24-Sep-1999

C:Accession: J05273  
R:Yokoyama, M.; Nishi, Y.; Yoshii, J.; Okubo, K.; Matsubara, K.  
DNA Res. 3, 311-320, 1996  
A:Title: Identification and cloning of neuroblastoma-specific and nerve tissue-specific  
A:Reference number: J05272; M01D:97191543  
A:Contents: neuroblastoma cell  
A:Accession: J05273  
A:status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-314 <OK>  
A:Cross-references: DDBJ:D82344; NID:q1841337; PIDD:BA011555.1; PID:d1012222; PID:q1841337  
A:Comment: This protein is a transcriptional repressor involved in regulating gene expression  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:99-155/Domain: homeobox homology <HOK>

RESULT 15  
T33110  
Hypothetical protein C18H7.3 - *Caenorhabditis elegans*  
C.Species: *Caenorhabditis elegans*  
C.Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000  
C.Accession: T33110  
R.Tin-Mollam, A.; Fironick, W.  
Submitted to the EMBL Data Library, May 1998  
A.Description: The sequence of C. elegans cosmid C18H7.  
A.Reference number: 221284  
A.Accession: T33110  
A.Status: preliminary; translated from GR/EMBL/DDbJ  
A.Molecule type: DNA  
A.Residues: 1-460 <TIN>  
A.Cross-references: EMBL:AF067607; PIDN:AAC17641.1; GSPDB:GN00022; CESP:C18H7.3  
A.Experimental source: strain Bristol N2; clone C18H7  
C.Genetics:  
A.Gene: CESP:C18H7.3  
A.Map position: 4  
A.Introns: 84/1  
A.Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Search completed: July 1, 2002, 06:28:10  
Job time: 712 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 1, 2002, 06:31:33 ; Search time 25.37 Seconds  
(without alignments)  
38.155 Million cell updates/sec

Title: US-09-461-684-2  
Perfect score: 109  
Sequence: 1 CEAAAAPAAAPAAAPAAAPAAAPAA 25

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt.40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75.5	69.3	518	1	TPM4_DROME
2	72	66.1	165	1	RS16_CAUCR
3	70	64.2	183	1	OLEC_BRANA
4	70	64.2	401	1	HME1_MOUSE
5	69	63.3	97	1	ANP_LIME
6	69	63.3	1028	1	OVO_DROME
7	68	62.4	673	1	HR38_DROME
8	68	62.4	1073	1	HR38_DROME
9	67	61.5	314	1	PMXR_MOUSE
10	67	61.5	314	1	PMXR_MOUSE
11	67	61.5	477	1	MA2_HUMAN
12	67	61.5	606	1	HMD_DROAN
13	66	60.6	366	1	HXAD_MOUSE
14	66	60.6	388	1	HXAD_MOUSE
15	65	59.6	85	1	ANP4_PSEAM
16	65	59.6	91	1	ANP4_PSEAM
17	65	59.6	276	1	SNX2_HUMAN
18	65	59.6	475	1	EVX2_MOUSE
19	65	59.6	476	1	EVX2_HUMAN
20	65	59.6	644	1	BT2_DROME
21	64	58.7	109	1	RLAI_TTCR
22	64	58.7	392	1	HME1_HUMAN
23	64	58.7	1533	1	PUM_DROME
24	64	58.7	2038	1	FSH_DROME
25	63	57.8	376	1	FXE1_HUMAN
26	63	57.8	521	1	RUN2_HUMAN
27	63	57.8	590	1	HMAA_DROME
28	63	57.8	607	1	RUN2_MOUSE
29	63	57.8	1095	1	PIPA_MOUSE
30	62	56.9	364	1	HK61_MESAU
31	62	56.9	365	1	HK61_RAT
32	62	56.9	1355	1	SALM_DROME
33	61	56.0	31	1	ANP3_PAGBO

## ALIGNMENTS

RESULT	ID	TPM4_DROME	STANDARD	PRT	518 AA.
34	61	56.0	91	1	ANP4_PSEAM
35	61	56.0	308	1	AEF1_DROME
36	61	56.0	376	1	FXI2_HUMAN
37	61	56.0	421	1	BR3A_MOUSE
38	61	56.0	423	1	BR3A_HUMAN
39	61	56.0	448	1	SPG1_STRSP
40	61	56.0	593	1	SPG2_STRSP
41	61	56.0	1733	1	VNDA_PRIVK
42	60	55.0	74	1	SR14_MACRA
43	60	55.0	82	1	ANP4_PSEAM
44	60	55.0	280	1	MACS_CHICK
45	60	55.0	367	1	HK61_HUMAN
1	TPM4_DROME	STANDARD	PRT	518 AA.	
AC	P49455:				
DT	01-FEB-1996 (Rel. 33, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	Tropomyosin I, fusion protein 33.				
GN	TMI OR TMI.				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
OC	Ephydroidea; Drosophilidae; Drosophila.				
OX	NCBI_TaxID=7227;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RP	MEDLINE=89127197; PubMed=2851721;				
BA	Hanke P.D., Scott R.V.;				
RT	"The drosophila melanogaster tropomyosin II gene produces multiple				
RT	proteins by use of alternative tissue-specific promoters and				
RT	alternative splicing."				
RL	Mol. Cell. Biol. 8:3591-3602(1988).				
RN	[2]				
RP	SEQUENCE OF 1-286 FROM N.A.				
RP	MEDLINE=87064486; PubMed=3097506;				
BA	Karlik C.C., Pyberg E.A.;				
RT	"Two Drosophila melanogaster tropomyosin genes: structural and				
RT	functional aspects."				
RL	Mol. Cell. Biol. 6:1965-1973(1986).				
CC	-1- ALTERNATIVE PRODUCTS: DROSOPHILA TROPOMYOSIN I GENE CAN PRODUCE				
CC	FOUR DIFFERENT ISOFORMS BY ALTERNATIVE PRODUCTS: A MUSCLE FORM,				
CC	A NON-MUSCLE FORM, AND TWO FUSION PROTEINS (33 AND 34).				
CC	-1- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE				
CC	EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.				
CC	-1- SIMILARITY: BELONGS TO THE TROPOMYOSIN FAMILY.				
CC	-1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN MAINLY BECAUSE				
CC	OF ALTERNATIVE EXON USAGE.				
CC	-----				
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	-----				
DR	EMBL: X76208; CAA53800.1; -				P23699 pseudopleur
DR	EMBL: L00362; AAA28965.1; -				P39413 drosophila
DR	EMBL: M12840; AAA28965.1; JOINED.				P58012 homo sapien
DR	EMBL: L00365; AAA28965.1; JOINED.				P17208 mus musculu
DR	EMBL: L00366; AAA28965.1; JOINED.				Q01851 homo sapien
DR	EMBL: L00367; AAA28965.1; JOINED.				P06654 streptococ
DR	EMBL: L00368; AAA28965.1; JOINED.				P19909 streptococ
DR	EMBL: L00369; AAA28965.1; JOINED.				P33485 pseudorabie
DR	HSSP: P04002; IATF.				O18881 macaca radi
DR	FLYBASE: FBgn0003721; Tml.				P04002 pseudopleur
DR					P16527 gallus gall
DR					P78426 homo sapien

```

DR InterPro: IPR000533; Tropomyosin.
DR Pfam: PF00261; Tropomyosin; 1.
DR PRINTS: PR00194; TROPOMYOSIN.
DR PROSITE: PS00326; TROPOMYOSIN; 1.
FT COILED COIL; Repeat; Alternative splicing; Multigene family.
FT DOMAIN 14 267 COILED COIL (POTENTIAL).
FT DOMAIN 287 518 ALA/PRO-RICH.
FT CONFLICT 106 114 LGSATKLS -> SASAIOLA (IN REF. 2).
FT CONFLICT 119 119 A -> S (IN REF. 2).
FT CONFLICT 183 183 A -> AMEADLERASERA (IN REF. 2).
FT CONFLICT 199 199 V -> L (IN REF. 2).
FT CONFLICT 215 231 MOREEYKNOIKTLNTR -> TQKEEFETQIKVLDS
(IN REF. 2).
SQ SEQUENCE 518 AA; 54558 MW; 153D0872CF9DB8EA CRC64;

Query Match 69.3%; Score 75.5; DB 1; Length 518;
Best Local Similarity 87.3%; Pred. No. 0.2;
Matches 21; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

OY 2 EAAAAEAAAAAAAAAAAAAAAA 25
DB 314 EAAAAEAAAAAAAAAEAAAAA 336

RESULT 2
RS16 CAUCR STANDARD; PRT; 165 AA.
ID RS16 CAUCR STANDARD; PRT; 165 AA.
AC P58122;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 30S ribosomal protein S16.
GN RPS16 OR CC3652.
OS Caulobacter crescentus.
CC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
CC Caulobacter.
CC NCBI_TaxID:63394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:ATCC 19089 / CB15;
RX MEDLINE:21173698; PubMed:11259647;
RA Nieman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Madcock J.R.,
RA Potocky I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeJoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Usterback T., Tran K., Wolf A., Yamathavan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RA "Complete genome sequence of Caulobacter crescentus.";
RA Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
RL -1 SIMILARITY: BELONGS TO THE S16P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE006023; AAK25614.1; -.
CC TIGR: CG3652; -.
DR InterPro: IPR000307; Ribosomal_S16.
DR Pfam: PF00886; Ribosomal_S16; 1.
DR PROSITE: PS00732; RIBOSOMAL_S16; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 165 AA; 17605 MW; ED46FC2798C5BE1C CRC64;

Query Match 66.1%; Score 72; DB 1; Length 165;
Best Local Similarity 75.0%; Pred. No. 0.18;
Matches 18; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 2 EAAAAEAAAAAAAAAAAAAAAA 25
DB 115 QAEADAKAAAEAKKAAAEAAAA 138

RESULT 3
OLEC_BRANA STANDARD; PRT; 183 AA.
ID OLEC_BRANA STANDARD; PRT; 183 AA.
AC P29526;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Oleosin C98 (Fragment).
GN C98
OS Brassica napus (rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Anther;
RX MEDLINE=93386188; PubMed=8374615;
RA Roberts M.R., Hodge R., Ross J.H.E., Sorensen A., Murphy D.J.,
RA Draper J., Scott R.;
RT "Characterization of a new class of oleosins suggests a male
RT gametophyte-specific lipid storage pathway.";
RT Plant J. 3:628-636(1993).
CC -1- FUNCTION: MAY HAVE A STRUCTURAL ROLE TO STABILIZE THE LIPID BODY
CC DURING DESICCATION OF THE SEED BY PREVENTING COALESCENCE OF THE
CC OIL. PROBABLY INTERACTS WITH BOTH LIPID AND PHOSPHOLIPID MOETIES
CC OF LIPID BODIES. MAY ALSO PROVIDE RECOGNITION SIGNALS FOR SPECIFIC
CC LIPASE ANCHORAGE IN LIPOLYSIS DURING SEEDLING GROWTH.
CC -1- SUBCELLULAR LOCATION: SURFACE OF OIL BODIES. OLEOSINS EXIST AT A
CC MONOLAYER LIPID/WATER INTERFACE.
CC -1- TISSUE SPECIFICITY: SPECIFIC TO THE MALE GAMETOPHYTE.
CC -1- SIMILARITY: BELONGS TO THE OLEOSIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X67142; CAA47623.1; -.
CC PIR: S24960; S24960.
DR HSP: P04002; 1MFA.
DR InterPro: IPR00136; Oleosin.
DR Pfam: PF01277; Oleosin; 1.
DR PROSITE: PS00811; OLEOSINS; 1.
KW Seed; Oil body; Multigene family.
FT NON_TER 1 1
FT DOMAIN <1 23 POLAR.
FT DOMAIN 24 95 HYDROPHOBIC.
SQ SEQUENCE 183 AA; 18149 MW; 198A5D3BBD3045A CRC64;

Query Match 64.2%; Score 70; DB 1; Length 183;
Best Local Similarity 78.3%; Pred. No. 0.31;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 AAAAAEAAAAAAAAAAAAAAAA 25
DB 151 AAPAAEPAPAAEAPAAEAPAA 173

RESULT 4
HME1_MOUSE STANDARD; PRT; 401 AA.
ID HME1_MOUSE STANDARD; PRT; 401 AA.
AC P09065;

```

DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Homeobox protein engrailed-1 (Mo-En-1).  
 GN EN1 OR EN-1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=93185339; PubMed=1363401;  
 RA Logan C., Hanks M.C., Noble-Topham S., Nallainathan D.,  
 RA Provart N.J., Joyner A.L.;  
 RT "Cloning and sequence comparison of the mouse, human, and chicken  
 RT engrailed genes reveal potential functional domains and regulatory  
 RT regions."; Dev. Genet. 13:345-356(1992).  
 RL Dev. Genet. 13:345-356(1992).  
 RN [2]  
 RP SEQUENCE OF 278-401 FROM N.A.  
 RX MEDLINE=88112776; PubMed=2892757;  
 RA Joyner A.L., Martin G.R.;  
 RT "En-1 and En-2, two mouse genes with sequence homology to the  
 RT Drosophila engrailed gene: expression during embryogenesis."; Genes  
 RL Genes Dev. 1:28-38(1987).  
 RN [3]  
 RP SEQUENCE OF 298-401 FROM N.A.  
 RX MEDLINE=86079501; PubMed=2416459;  
 RA Joyner A.L., Kornberg T., Coleman K.G., Cox D.R., Martin G.R.;  
 RT "Expression during embryogenesis of a mouse gene with sequence  
 RT homology to the Drosophila engrailed gene."; Cell 43:29-37(1985).  
 RL Cell 43:29-37(1985).  
 RN [4]  
 RP SEQUENCE OF 321-380 FROM N.A.  
 RX MEDLINE=91095509; PubMed=1980115;  
 RA Holland P.W.H., Williams N.A.;  
 RT "Conservation of engrailed-like homeobox sequences during vertebrate  
 RT evolution"; FEBS Lett. 277:250-252(1990).  
 RL FEBS Lett. 277:250-252(1990).  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: BELONGS TO THE ENGRAILED FAMILY OF HOMEBOX PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L12703; AAA03660.2; -  
 DR EMBL: Y00201; CA68361.1; -  
 DR PIR: A26629; A26629.  
 DR PIR: A24778; A24778.  
 DR PIR: S13009; S13009.  
 DR PIR: A48423; A48423.  
 DR HSP: P02836; 3HDD.  
 DR TRANSFAC: T02016; -  
 DR MGD: MGI:95389; En1.  
 DR InterPro: IPR000747; Engrailed.  
 DR InterPro: IPR001356; Homeobox.  
 DR Pfam: PF00046; homeobox; 1.  
 DR PRINTS: PR00026; ENGRAILED.  
 DR SMART: SM00389; HOX; 1.  
 DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 DR PROSITE: PS00071; HOMEBOX\_2; 1.  
 DR PROSITE: PS00033; ENGRAILED; 1.  
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.  
 FT DOMAIN 52 87 PRO-RICH.  
 FT DOMAIN 73 87 POLY-PRO.  
 FT DOMAIN 207 228 POLY-ALA.  
 FT DNA\_BIND 312 371 HOMEBOX.

SEQ SEQUENCE 401 AA: 40950 MW: 1F90210950152FAE CRC64;  
 Query Match 64.2%; Score 70; DB 1; Length 401;  
 Best Local Similarity 78.3%; Pred. No. 0.57;  
 Matches 18; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 3 AAAAAEAAAAAEEAAAAA 25  
 Db 207 AAAAAAATAAAAAA 229  
 RESULT 5  
 ANP\_LIMFE STANDARD: PRT: 97 AA.  
 ID ANP\_LIMFE  
 AC P09031.  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DE Antifreeze protein precursor (AFP).  
 DE Limanda ferruginea (Yellowtail flounder).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 CC Acanthomorpha; Acanthopterygii; Perciformes; Pleuronectiformes;  
 CC Pleuronectoidae; Pleuronectidae; Limanda.  
 NX NCBI\_TaxID=8258;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88029483; PubMed=365937;  
 RA Scott G.K., Davies P.L., Shears N.A., Fletcher G.L.;  
 RT "Structural variations in the alanine-rich antifreeze proteins of the  
 RT pleuronectinae"; Eur. J. Biochem. 168:629-633(1987).  
 RL Eur. J. Biochem. 168:629-633(1987).  
 CC -1- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.  
 CC -1- SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE 1 AFP ARE  
 CC ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X06356; CAA29655.1; -  
 DR PIR: S02376; S02376.  
 DR HSP: P04002; 1ATF.  
 DR InterPro: IPR000104; Antifreeze\_1.  
 DR PRINTS: PR00308; ANTIFREEZE1.  
 KW Antifreeze protein; Repeat; Signal.  
 FT SIGNAL 1 23  
 FT PROPEP 24 48  
 FT CHAIN 49 97  
 FT SEQUENCE 97 AA: 8865 MW: 62AD582DF8E459B6 CRO64;  
 Query Match 63.3%; Score 69; DB 1; Length 97;  
 Best Local Similarity 70.8%; Pred. No. 0.24;  
 Matches 17; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
 QY 2 EAAAAEAAAAAEEAAAAA 25  
 Db 53 DAAAAAATAAAAAAADTAAA 76  
 RESULT 6  
 OVO\_DROME STANDARD: PRT: 1028 AA.  
 ID OVO\_DROME  
 AC P51521; Q9XZU4;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)

16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ovo protein (Shaven baby protein).  
 GN OVO OR SVB.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_Taxid=7227;  
 RX MEDLINE=95021209; PubMed=7935398;  
 RA Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;  
 RT "Multiple products from the shavenbaby-ovo gene region of Drosophila  
 melanogaster: relationship to genetic complexity";  
 RL Mol. Cell. Biol. 14:6809-6818(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OREGON-R;  
 RX MEDLINE=91293102; PubMed=1712294;  
 RA Meyel-Nitlo M.T.M., Terracoli R., Kafatos F.C.;  
 RT "The ovo gene of Drosophila encodes a zinc finger protein required  
 for female germ line development.";  
 RL EMBO J. 10:2259-2266(1991).  
 CC -1- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMINUM AND  
 CC ACCUMULATES IN NURSE CELLS DURING OOGENESIS. STORED IN THE EGG,  
 CC BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED  
 CC PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U11383; AAB60216.1; -;  
 DR EMBL: X59772; CAB36921.1; ALT\_SRD.  
 DR HSSP: P25490; 12NM.  
 DR FlyBase; FBgn003028; ovo.  
 DR InterPro; IPR000822; znf-C2H2.  
 DR Pfam; PF00096; zf-C2H2; 4.  
 DR PRINTS; PR00048; ZINC\_FINGER.  
 DR SMART; SM00355; znf\_C2H2; 4.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 3.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 3.  
 KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;  
 KW Transcription regulation.  
 FT DOMAIN 62 66 POLY-ALA.  
 FT DOMAIN 72 77 POLY-GLY.  
 FT DOMAIN 80 85 POLY-GLY.  
 FT DOMAIN 98 108 POLY-GLY.  
 FT DOMAIN 144 152 POLY-HIS.  
 FT DOMAIN 153 159 POLY-ASN.  
 FT DOMAIN 336 339 POLY-GLN.  
 FT DOMAIN 347 353 POLY-GLN.  
 FT DOMAIN 357 361 POLY-GLN.  
 FT DOMAIN 410 414 POLY-GLN.  
 FT DOMAIN 418 422 POLY-GLN.  
 FT DOMAIN 426 432 POLY-GLN.  
 FT DOMAIN 432 433 POLY-GLN.  
 FT DOMAIN 445 453 POLY-GLN.  
 FT DOMAIN 456 459 POLY-GLN.  
 FT DOMAIN 466 474 POLY-GLN.  
 FT DOMAIN 497 517 POLY-ALA.  
 FT DOMAIN 524 529 POLY-SER.  
 FT DOMAIN 539 558 POLY-ALA.  
 FT DOMAIN 639 651 POLY-ALA.  
 FT DOMAIN 717 725 POLY-ALA.  
 FT DOMAIN 797 802 POLY-GLN.

FT DOMAIN 820 823 POLY-GLN.  
 FT DOMAIN 826 832 POLY-GLN.  
 FT DOMAIN 874 992 ZINC FINGERS.  
 FT ZN\_FING 874 896 C2H2-TYPE.  
 FT ZN\_FING 902 924 C2H2-TYPE.  
 FT ZN\_FING 930 953 C2H2-TYPE.  
 FT ZN\_FING 969 992 C2H2-TYPE.  
 FT CONFLICT 647 647 A -> R (IN REF. 2).  
 SQ SEQUENCE 1028 AA; 110620 MW; D7058B2BC0F6F77 CRC64;  
 Query Match 63.3%; Score 69; DB 1; Length 1028;  
 Best Local Similarity 85.7%; Pred. No. 1.5;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 3 AAAAAAAAAAAAAAAAAAAAAA 23  
 DB 497 AAAAAAAAAAAAAAAAAAAAAA 517  
 SIM\_DROME 7  
 ID SIM\_DROME STANDARD; PRT; 673 AA.  
 AC P05709; O96521; Q9VF23;  
 DT 01-NOV-1988 (Rel. 22, Created)  
 DI 01-MAY-1992 (Rel. 22, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Single-minded protein.  
 GN SIM OR CG7771.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_Taxid=7227;  
 RX MEDLINE=8151023; PubMed=3345560;  
 RX Crews S.T., Thomas J.B., Goodman C.S.;  
 RT "The Drosophila single-minded gene encodes a nuclear protein with  
 RT sequence similarity to the per gene product";  
 RL Cell 52:143-151(1988).  
 RN [2]  
 RP SEQUENCE OF 1-18 FROM N.A. AND SIMILARITY TO HLH PROTEINS.  
 RX MEDLINE=92103681; PubMed=1760843;  
 RA Nambu J.R., Lewis J.O., Wharton K.A. Jr., Crews S.T.;  
 RT "The Drosophila single-minded gene encodes a helix-loop-helix protein  
 RT that acts as a master regulator of CNS midline development.";  
 RL Cell 67:1157-1167(1991).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99054545; PubMed=9840810;  
 RA Kasal Y., Stahl S., Crews S.;  
 RT "Specification of the Drosophila CNS midline cell lineage: direct  
 RT control of single-minded transcription by dorsal/ventral patterning  
 RT genes";  
 RL Gene Expr. 7:171-189(1998).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Scherker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherker S.E., Li P.W., Hoskins R., Henderson S.N.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Abmayyan A., An H.-J., Andrews-Pfankoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Bessu P.V., Bertan B.P., Bandaru D., Bolstad S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brattler P.,  
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,



RA Dodson K., Dopp L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foslø C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glöckler A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibeagwam C.,  
 RA Jalali M., Kalush F., Karpén G.H., Ke Z., Kennison J.A., Kerchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Krevitz S., Kulp D., Lai Z.,  
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Meriklov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphay L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Sigen Klamas I., Simpson M., Skupski M.F., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RL "The genome sequence of *Drosophila melanogaster*.";  
 Science 287:2185-2195(2000).  
 CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT FUNCTIONS AS A MASTER  
 CC DEVELOPMENTAL REGULATOR OF THE CNS MIDLINE LINEAGE. MUTATIONS IN  
 CC THE SIM GENE RESULTS IN THE LOSS OF THE PRECURSOR CELLS GIVING  
 CC RISE TO MIDLINE CELLS OF THE EMBRYONIC CENTRAL NERVOUS SYSTEM.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS.  
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib.ch](mailto:license@sib.ch)).  
 CC -----  
 DR EMBL: M19020; AA28900.1; -  
 DR EMBL: AF071934; AAC64519.1; ALT\_SEQ.  
 DR EMBL: AE003698; AAF54902.1; ALT\_SEQ.  
 DR PIR: A29945; A29945.  
 DR TRANSFAC: T00750; -  
 DR FlyBase: FBgn0004666; sim.  
 DR InterPro: IPR003015; HLH\_MYC.  
 DR InterPro: IPR001092; HLH\_dlm.  
 DR InterPro: IPR001610; PAC.  
 DR InterPro: IPR000014; PAS.  
 DR Pfam: PF00785; PAC; 1.  
 DR Pfam: PF00989; PAS; 2.  
 DR SMART: SM00353; HLH; 1.  
 DR SMART: SM00086; PAC; 1.  
 DR SMART: SM00091; PAS; 2.  
 DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.  
 DR PROSITE: PS0112; PAS; 2.  
 KW Developmental protein; Neurogenesis; Nuclear protein; Repeat;  
 KW Transcription regulation; DNA-binding  
 FT DNA\_BIND 1 13  
 FT DOMAIN 14 13  
 FT DOMAIN 54 54  
 FT DOMAIN 76 148  
 FT DOMAIN 242 312  
 FT DOMAIN 382 422  
 FT DOMAIN 423 426  
 FT DOMAIN 484 488  
 FT DOMAIN 489 489  
 FT DOMAIN 496 499  
 FT DOMAIN 501 505  
 FT POLY-GLN.

FT DOMAIN 529 533  
 FT DOMAIN 573 576  
 FT DOMAIN 604 607  
 FT DOMAIN 649 669  
 FT CONFLICT 127 127  
 FT CONFLICT 401 409  
 SC SEQUENCE 673 AA; 73589 MW; 29F0ABBA2BC0EBC CRC64;  
 Query Match 62.4%; Score 68; DB 1; Length 673;  
 Best Local Similarity 66.7%; Pred. No. 1.3;  
 Matches 16; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 Y 2 EAAAEAAAEAAAEAAAEAAAEAA 25  
 Db 384 QAAQAAQAAQAAQAAQAAQAA 407  
 RESULT 8  
 ID HR38\_DROME STANDARD; PRT; 1073 AA.  
 AC P49868; O18383; OSYTKA;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Probable nuclear hormone receptor HR38 (dHR38).  
 GN HR38 OR NR4A4 OR CG1864.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).  
 RC TISSUE=Larva;  
 RX MEDLINE=95372400; PubMed=7644522;  
 RA Sutherland J.D., Kozlova T., Tzertzinis G., Kafatos F.C.;  
 RT "Drosophila hormone receptor 38: a second partner for Drosophila USP  
 RT suggests an unexpected role for nuclear receptors of the nerve growth  
 RT factor-induced protein B type.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:7966-7970(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A. (SHORT ISOFORM), AND TISSUE SPECIFICITY.  
 RX MEDLINE=96370123; PubMed=9704500;  
 RA Komonyi O., Mink M., Csiba U., Maroy P.;  
 RT "Genomic organization of DHR38 gene in *Drosophila*: presence of  
 RT Alu-like repeat in a translated exon and expression during embryonic  
 RT development.";  
 RL Arch. Insect Biochem. Physiol. 38:185-192(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A. (LONG ISOFORM), AND CHARACTERIZATION.  
 RC TISSUE=Larva;  
 RX MEDLINE=98315108; PubMed=9649534;  
 RA Kozlova T., Pokholkova G.V., Tzertzinis G., Sutherland J.D.,  
 RA Zhimulev I.F., Kafatos F.C.;  
 RT "Drosophila hormone receptor 38 functions in metamorphosis: a role in  
 RT adult cuticle formation.";  
 RL Genetics 149:1465-1475(1998).  
 RN [4]  
 RP SEQUENCE FROM N.A. (LONG ISOFORM).  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Bayendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,

DR PRINTS: PR000398; STRDHOMRNER.  
DR PRINTS: PR000447; STRDIOFINGER.  
DR SMART: SM00430; HOL1. 1.  
DR SMART: SM00399; ZNF\_C4. 1.  
DR PROSITE: PS00031; NUCLEAR\_RECEPTOR; 1.  
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
KW Zinc-finger; Alternative splicing; Developmental protein.  
FT DNA\_BIND 744 809 NUCLEAR RECEPTOR-TYPE.  
FT ZN\_FING 780 804 C4-TYPE.  
FT DOMAIN 188 192 POLY-ALA.  
FT DOMAIN 206 218 POLY-ALA.  
FT DOMAIN 221 228 POLY-ALA.  
FT DOMAIN 268 272 POLY-THR.  
FT DOMAIN 294 312 POLY-GLN.  
FT DOMAIN 441 462 POLY-GLN.  
FT DOMAIN 505 508 POLY-SER.  
FT DOMAIN 619 626 POLY-GLN.  
FT DOMAIN 661 665 POLY-ALA.  
FT VARSPLIT 1 522 MISSING (IN SHORT ISOFORM).  
FT CONFLICT 667 667 V -> VSPSP (IN REF. 4).  
FT CONFLICT 685 682 S -> L (IN REF. 1 AND 3).  
FT CONFLICT 689 692 STAG -> LGSR (IN REF. 2).  
FT CONFLICT 697 697 A -> D (IN REF. 2).  
FT CONFLICT 702 702 N -> S (IN REF. 2).  
FT CONFLICT 1041 1041 S -> R (IN REF. 2).  
FT CONFLICT 1064 1064 E -> D (IN REF. 2).  
SQ SEQUENCE 1073 AA; 116991 MW; 126A30DAFA1C096A CRC64;

Query Match  
Best Local Similarity 73.9%; Score 68; DB 1; Length 1073;  
Matches 1/; Conservative 0; Pred. No. 1.9;  
Mismatches 6; Indels 0; Gaps 0;

3 AAAAAAAAAAAAAAAAAAAAAA 25  
Db 206 AATAATATAAEAGGASAAAAA 228

RESULT 9  
PMXB\_HUMAN ID ID PMXB\_HUMAN STANDARD; PRT; 314 AA.  
AC 099453;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Paired mesoderm homeobox protein 2B (Paired-like homeobox 2B)  
DE (PHOX2B homeodomain protein) (Neuroblastoma Phox) (NBPhox).  
GN PMX2B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RP TISSUE=Neuroblastoma;  
RX MEDLINE=97191543; PubMed=9039501;  
RX MEDLINE=97191543; PubMed=9039501;  
RA Yokoyama M., Nishi Y., Yoshii J., Okubo K., Matsubara K.;  
RT "Identification and cloning of neuroblastoma-specific and nerve  
RT tissue-specific genes through compiled expression profiles.";  
RL DNA Res. 3:311-320(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RP MEDLINE=20487360; PubMed=11034547;  
RX Adachi M., Browne D., Lewis E.J.;  
RT "Paired-like homeodomain proteins Phox2a/Arx and Phox2b/NBPhox have  
RT beta-hydroxylase gene organization and independently regulate dopamine  
RL DNA Cell Biol. 19:539-554(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99326521; PubMed=10395796;  
RX Yokoyama M., Watanabe H., Nakamura M.;

RT "Genomic structure and functional characterization of NBPHox (PMX2B),  
 RT a homeodomain protein specific to catecholaminergic cells that is  
 RT involved in second messenger-mediated transcriptional activation."  
 RL Genomics 59:40-50(1999)  
 CC -1- FUNCTION: INVOLVED IN THE DEVELOPMENT OF SEVERAL MAJOR  
 CC NORADRENERGIC NEURON POPULATIONS, INCLUDING THE LOCUS COERULEUS.  
 CC TRANSCRIPTION FACTOR WHICH COULD DETERMINE A NEURORANSMITTER  
 CC PHENOTYPE IN VERTEBRATES. ENHANCES SECOND-MESSENGER-MEDIATED  
 CC ACTIVATION OF THE DOPAMINE BETA-HYDROLASE AND C-FOS PROMOTERS,  
 CC AND OF SEVERAL ENHANCERS INCLUDING CYCLIC AMP-RESPONSE ELEMENT AND  
 CC SERUM-RESPONSE ELEMENT.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN NEUROBLASTOMA, BRAIN AND ADRENAL  
 CC GLAND.  
 CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D82344; BA01555.1; -  
 DR EMBL: AF117379; A026698.1; -  
 DR EMBL: AB015671; BA082670.1; -  
 DR HSSP: P14653; 1B72.  
 DR MIM: 603851; -  
 DR InterPro: IPR001356; Homeobox.  
 DR Pfam: PF00046; homeobox; 1.  
 DR PRINTS: PR00024; HOMEBOX.  
 DR SMART: SM00389; HOX; 1.  
 DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 DR PROSITE: PS50071; HOMEBOX\_2; 1.  
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 KW Transcription regulation.  
 FT DNA\_BIND 98 157 HOMEBOX.  
 FT DOMAIN 159 167 POLY-ALA.  
 FT DOMAIN 212 217 POLY-GLY.  
 FT DOMAIN 241 260 POLY-ALA.  
 FT SEQUENCE 314 AA; 31607 MW; 76737F71948B5D81 CRC64;  
 SQ  
 Query Match 61.5%; Score 67; DB 1; Length 314;  
 Best Local Similarity 78.3%; Pred. No. 0.95;  
 Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 3 AAAAAEAAAAEAAAAEAAAAA 25  
 Db 244 AAAAAAAAAAAAAAAAAAGLAAA 266  
 RESULT 10  
 ID PMXB\_MOUSE STANDARD; PRT; 314 AA.  
 AC 035690;  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Paired mesoderm homeobox protein 2B (Paired-like homeobox 2B)  
 DE (PHOX2B homeodomain protein) (Neuroblastoma Phox) (NBPhox).  
 OS Mus musculus (Mouse).  
 GN Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=98040559; PubMed=9374403;  
 RA Patten A., Morin X., Cremer H., Goridis C., Brunet J.-F.;  
 RT "Expression and interactions of the two closely related homeobox  
 RT genes Phox2a and Phox2b during neurogenesis.";

RL Development 124:4065-4075(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99326521; PubMed=10395798;  
 RX Yokoyama M., Watanabe H., Nakamura M.;  
 RT "Genomic structure and functional characterization of NBPHox (PMX2B),  
 RT a homeodomain protein specific to catecholaminergic cells that is  
 RT involved in second messenger-mediated transcriptional activation."  
 RL Genomics 59:40-50(1999).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Y14493; CA07483.1; -  
 DR EMBL: AB015672; BA082671.1; -  
 DR HSSP: P06601; 1F3L.  
 DR MGD: MGI:1100882; Pmx2b.  
 DR InterPro: IPR00047; HTH\_repressr.  
 DR InterPro: IPR001356; Homeobox.  
 DR Pfam: PF00046; homeobox; 1.  
 DR PRINTS: PR00024; HOMEBOX.  
 DR PRINTS: PR00031; HTHREPRESSR.  
 DR SMART: SM00389; HOX; 1.  
 DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 DR PROSITE: PS50071; HOMEBOX\_2; 1.  
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.  
 FT DNA\_BIND 98 157 HOMEBOX.  
 FT DOMAIN 159 167 POLY-ALA.  
 FT DOMAIN 212 217 POLY-GLY.  
 FT DOMAIN 241 260 POLY-ALA.  
 FT SEQUENCE 314 AA; 31621 MW; 40737F71948B595A CRC64;  
 SQ  
 Query Match 61.5%; Score 67; DB 1; Length 314;  
 Best Local Similarity 78.3%; Pred. No. 0.95;  
 Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 3 AAAAAEAAAAEAAAAEAAAAA 25  
 Db 244 AAAAAAAAAAAAAAAAAAGLAAA 266  
 RESULT 11  
 ID MAZ\_HUMAN STANDARD; PRT; 477 AA.  
 AC P56270; Q99443; Q15703;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE MYC-associated zinc finger protein (MAZ1) (Purine-binding  
 DE transcription factor) (Pur-1) (Zf87) (Zf87).  
 GN MAZ.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=92366479; PubMed=1502157;  
 RA Bessone S.A., Asselin C., Patel A.J., Marcu K.B.;  
 RT "MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences  
 RT regulating transcriptional initiation and termination."  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:7452-7456(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Carcinoma;

RX MEDLINE=9232709; PubMed=1567856;  
 RA Pyro J.J., Moberg K.H., Hall D.J.;  
 RT "Isolation of a novel cDNA encoding a zinc-finger protein that binds  
 RT to two sites within the c-myc promoter.";   
 RL Biochemistry 31:4102-4110(1992).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreatic Islets;  
 RA MEDLINE=96428591; PubMed=8831693;  
 RA Tsuboi H., Sakatsune O., Itakura K., Yokoyama K.K.;  
 RT "Members of the MAZ family: a novel cDNA clone for MAZ from human  
 RT pancreatic islet cells.";   
 RL Biochem. Biophys. Res. Commun. 226:801-809(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=96224025; PubMed=8626793;  
 RA Parks C.L., Shenk T.;  
 RT "The serotonin 1a receptor gene contains a TATA-less promoter that  
 RT responds to MAZ and Sp1.";   
 RL J. Biol. Chem. 271:4417-4430(1996).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lymphoblastoma;  
 RA MEDLINE=96352105; PubMed=9685418;  
 RA Song J., Murakami H., Tsuboi H., Tang X., Matsumura M., Itakura K.,  
 RA Kanazawa I., Sun K., Yokoyama K.K.;  
 RT "Genomic organization and expression of a human gene for Myc-  
 RT associated zinc finger protein (MAZ).";   
 RL J. Biol. Chem. 273:20603-20614(1998).  
 CC -1- FUNCTION: MAZ FUNCTION AS A TRANSCRIPTION FACTOR WITH DUAL ROLES  
 CC IN TRANSCRIPTION INITIATION AND TERMINATION. BINDS TO TWO SITES,  
 CC MEAL AND MEAL2, WITHIN THE C-MYC PROMOTER HAVING GREATER  
 CC AFFINITY FOR THE FORMER. ALSO BINDS TO MULTIPLE G/C-RICH SITES  
 CC WITHIN THE PROMOTER OF THE SPI FAMILY OF TRANSCRIPTION FACTORS.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- TISSUE SPECIFICITY: HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL  
 CC MUSCLE, AND PANCREAS. SEEMS NOT TO BE EXPRESSED IN KIDNEY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M94046; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; M93339; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; D85131; BA012728.1; ALT\_INIT.  
 DR EMBL; U3819; AAB04121.1; ALT\_INIT.  
 DR EMBL; AB017335; BAA33064.1; -;  
 DR HSSP; P08046; 1AAY.  
 DR TRANSFAC; T00490; -;  
 DR TRANSFAC; T02305; -;  
 DR MIM; 600999; -;  
 DR InterPro; IPR000822; Znf-C2H2.  
 DR Pfam; PF00096; zf-C2H2\_6.  
 DR PRINTS; PR00048; ZINC\_FINGER.  
 DR SMART; SM00355; Znf\_C2H2\_6.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 5.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 5.  
 KW Transcription regulation; zinc-finger; Metal-binding; DNA-binding;  
 KW RNA-binding; Repeat; Nuclear protein.  
 FT DOMAIN 190 413 ZINC\_FINGERS.  
 FT ZN\_FING 190 212 C2H2-TYPE.  
 FT ZN\_FING 279 301 C2H2-TYPE.  
 FT ZN\_FING 307 329 C2H2-TYPE.  
 FT ZN\_FING 337 360 C2H2-TYPE.  
 FT ZN\_FING 366 388 C2H2-TYPE.  
 FT ZN\_FING 392 413 C2H2-TYPE.  
 FT ZN\_FING 96 108 POLY-ALA.  
 FT DOMAIN 133 139 POLY-PRO.  
 FT DOMAIN 157 161 POLY-ALA.

FT DOMAIN 245 249 POLY-GLY.  
 FT DOMAIN 435 449 POLY-ALA.  
 FT CONFLICT 259 259 MISSING (IN REF. 3).  
 FT CONFLICT 401 401 L->M (IN REF. 2 AND 4).  
 FT CONFLICT 443 447 MISSING (IN REF. 3).  
 SQ SEQUENCE 477 AA; 48607 MW; C04C80F3C3C6825 CnC64;

Query Match 61.5%; Score 67; DB 1; Length 477;  
 Best Local Similarity 77.3%; Pred. No. 1.3;  
 Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 4 AAAAAAAAAAAAAAAAAAAAAA 25  
 DB 90 AAAAAAAAAAAAAAAAAAAAAA 111

RESULT 12  
 HMID\_DROAN STANDARD; PRT; 606 AA.  
 ID HMID\_DROAN  
 AC P22544;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Homeobox protein OM(1D).  
 GN OM(1D).  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 ON NCBI\_TaxId=7217;  
 RX MEDLINE=91122048; PubMed=1671353;  
 RA Tanga S., Cores V.G.;  
 RT "Retroransposon-induced overexpression of a homeobox gene causes  
 RT defects in eye morphogenesis in Drosophila.";   
 RL EMBO J. 10:407-417(1991).  
 CC -1- FUNCTION: Probably involved in eye morphogenesis.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEOBOX PROTEINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X56682; CAA40011.1; -;  
 DR PIR; S13367; S13367.  
 DR HSSP; P22808; 1VND.  
 DR Flybase; FBgn0012114; Dana\B.  
 DR InterPro; IPR001356; Homeobox.  
 DR Pfam; PF00046; homeobox\_1.  
 DR PRINTS; PR00024; HOMEOBOX.  
 DR SMART; SM00389; HOX\_1.  
 DR PROSITE; PS00027; HOMEOBOX\_1; 1.  
 DR PROSITE; PS50071; HOMEOBOX\_2; 1.  
 KW DNA-binding; Homeobox; Developmental protein; Nuclear protein; Vision.  
 FT DOMAIN 23 57 HIS/GLN-RICH (OPA-REPEAT).  
 FT DOMAIN 106 124 HIS/GLN-RICH (OPA-REPEAT).  
 FT DOMAIN 173 193 HIS/PRO-RICH.  
 FT DNA\_BIND 331 390 HOMEOBOX.  
 FT DOMAIN 220 248 ALA-RICH.  
 FT DOMAIN 422 434 ALA-RICH.  
 FT DOMAIN 450 455 ALA-RICH.  
 FT DOMAIN 503 510 ALA-RICH.  
 FT DOMAIN 515 521 PRO-RICH.  
 SO SEQUENCE 606 AA; 61735 MW; AA7B8B6367370FBB CnC64;

Query Match 61.5%; Score 67; DB 1; Length 606;  
 Best Local Similarity 78.3%; Pred. No. 1.6;  
 Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AAAAAAAAAAAAAAAAAAAAAA 25  
 DB 220 AAAAAAAAAAAAAAAAAAAAAA 242

RESULT 13  
 HXAD\_MOUSE STANDARD; PRT; 386 AA.  
 AC Q62424;  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-OCT-2001 (Rel. 40, Last annotation update)  
 DE Homeobox protein Hox-A13 (Hox-1.10).  
 GN HOXA13 OR HOX-1.10.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96259555; PubMed=8673126;  
 RA Mortlock D.P., Post L.C., Innis J.W.;  
 RT "The molecular basis of hypodactyly (hd): a deletion in Hoxa 13 leads  
 RT to arrest of digital arch formation.";  
 RL Nat. Genet. 13:284-289(1996).  
 CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF  
 CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH  
 CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS (BY  
 CC SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- DISEASE: DEFECTS IN HOXA13 ARE THE CAUSE OF HYPODACTYLY (HD), A  
 CC CONDITION CHARACTERIZED BY PROFOUND DEFICIENCY OF DIGITAL ARCH  
 CC STRUCTURES.  
 CC -1- SIMILARITY: BELONGS TO THE ABD-B FAMILY OF HOMEBOX PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U59322; AAB03322.1; -  
 CC DR HSSP: P02833; ISAN.  
 CC DR TRANSFAC: T03337; -  
 CC DR MGD: MGI:96173; Hoxa13.  
 CC DR InterPro: IPR001356; Homeobox.  
 CC DR Pfam: PF00046; homeobox; 1.  
 CC DR SMART: SM00389; HOX; 1.  
 CC DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 CC DR PROSITE: PS50071; HOMEBOX\_2; 1.  
 CC KM Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 CC Transcription regulation.  
 CC FT DOMAIN 38 51 POLY-ALA.  
 CC FT DNA\_BIND 320 379 HOMEBOX.  
 CC FT DOMAIN 52 57 POLY-GLY.  
 CC FT DOMAIN 62 66 POLY-ALA.  
 CC FT DOMAIN 73 84 POLY-ALA.  
 CC FT DOMAIN 101 104 POLY-ALA.  
 CC FT DOMAIN 116 133 POLY-ALA.  
 CC FT DOMAIN 198 205 POLY-ALA.  
 CC SQ SEQUENCE 386 AA; 39566 MW; ZB01DCC9B1951324 CRC64;

Query Match 60.6%; Score 66; DB 1; Length 386;  
 Best Local Similarity 70.8%; Pred. No. 1.4;  
 Matches 17; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 EAAAAAAAAAAAAAAAAAAAAA 25  
 DB 111 EAPPSAAAAAAAAAAAAAAAAAAS 134

RESULT 14  
 HXAD\_HUMAN STANDARD; PRT; 388 AA.  
 AC P31271; Q63371;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Homeobox protein Hox-A13 (Hox-1J).  
 GN HOXA13 OR HOX1J.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97172976; PubMed=9020844;  
 RA Mortlock D.P., Innis J.W.;  
 RT "Mutation of HOXA13 in hand-foot-genital syndrome.";  
 RL Nat. Genet. 15:179-180(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Bradshaw H., Hinds K., Keppeler D.;  
 RL Submitted (Feb-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 322-387 FROM N.A.  
 RX MEDLINE=90098676; PubMed=2574852;  
 RA Acampora D., D'Esposito M., Falcetta A., Panese M., Migliaccio E.,  
 RA Morelli F., Stornaiuolo A., Nigro V., Simeone A., Boncinelli E.;  
 RT "The human Hox gene family.";  
 RL Nucleic Acids Res. 17:10385-10402(1989).  
 CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF  
 CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH  
 CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: BELONGS TO THE ABD-B FAMILY OF HOMEBOX PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U82827; AAC50993.1; -  
 CC DR EMBL: AC004080; -; NOT\_ANNOTATED\_CDS.  
 CC DR PIR: S14932; S14932.  
 CC DR HSSP: P02833; ISAN.  
 CC DR TRANSFAC: T03321; -  
 CC DR MIM: 142959; -  
 CC DR InterPro: IPR001356; Homeobox.  
 CC DR Pfam: PF00046; homeobox; 1.  
 CC DR SMART: SM00389; HOX; 1.  
 CC DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 CC DR PROSITE: PS50071; HOMEBOX\_2; 1.  
 CC KM Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 CC Transcription regulation.  
 CC FT DNA\_BIND 322 381 HOMEBOX.  
 CC FT DOMAIN 38 53 POLY-ALA.  
 CC FT DOMAIN 62 66 POLY-ALA.  
 CC FT DOMAIN 73 84 POLY-ALA.  
 CC FT DOMAIN 116 133 POLY-ALA.  
 CC FT DOMAIN 145 150 POLY-ALA.  
 CC FT DOMAIN 200 207 POLY-ALA.  
 CC FT CONFLICT 146 146 A -> G (IN REF. 2).  
 CC FT CONFLICT 147 147 A -> H (IN REF. 2).  
 CC FT CONFLICT 187 187 P -> A (IN REF. 2).  
 CC FT CONFLICT 195 195 P -> A (IN REF. 2).  
 CC FT CONFLICT 198 198 P -> A (IN REF. 2).

SO SEQUENCE 388 AA: 39752 MW: 6CD9C9A5616C2FE6 CRC64:

Query Match	60.6%	Score 66;	DB 1;	Length 388;
Best Local Similarity	70.8%;	Pred. No. 1.4;		
Matches 17; Conservative	2;	Mismatches 5;	Indels 0;	Gaps 0;

```

OY      2  EAAAAAAAAAAAAAAAAEAAAAA  25
          || :| ||||| ||||| ||||| :
Db      111 EAPPSAAAAAAAAAAAAAAAAAS 134

```

RESULT	15
ANP4_PSEAM	
ID	ANP4_PSEAM
STANDARD;	PRT;
	95 AA.

```

DT      21-JUL-1986 (Rel. 01, Created)
DT      13-AUG-1987 (Rel. 05, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Antifreeze peptide 4 precursor.
OS      Pseudopleuronectes americanus (Winter flounder) (Pleuronectes
OC      americanus).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC      Acanthomorphi; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC      Pleuronectoidae; Pleuronectidae; Pseudopleuronectes.
OC      NCBI_TaxId=8265;
RN      [1]
RX      SEQUENCE FROM N.A.
RX      MEDLINE=81247379; PubMed=6265915;
RA      Lin Y., Gross J.R.;
RT      "Molecular cloning and characterization of winter flounder antifreeze
RT      cDNA."
RL      Proc. Natl. Acad. Sci. U.S.A. 78:2825-2829(1981)
CC      FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
CC      SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE I AFP ARE
CC      ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.
CC      CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-stb.ch/announce/
CC      or send an email to license@isb-stb.ch).
CC      -----
DR      EMBL, J00930, AAA9467.1, -.
DR      PIR, A03193, FDEL4W.
DR      HSSP, P04002, 1WFA.
DR      InterPro, IPR000104, Antifreeze_1.
DR      PRINTS, PR00308, ANTIFREEZE1.
DR      KMW Antifreeze protein; Repeat; Multigene family; Signal.
FT      SIGNAL 1 21 85
FT      CHAIN 22 85 ANTIFREEZE PEPTIDE 4.
SEQUENCE 85 AA: 7215 MW: 8662EID2B44117BC CRC64:

```

Query Match	59.6%	Score 65;	DB 1;	Length 85;
Best Local Similarity	73.9%	Pred. NO. 0.55;		
Matches	17;	Conservative	0;	Mismatches 6;
			Indels	0;
			Gaps	0;

QY 3 AAAAAEAAAAEAAAAA 25  
 |||| | | |||| | |  
 Db 28 AAAAAATATAAAAAAATTAATAA 50

Search completed: July 1, 2002, 06:31:34  
Job time: 671 sec



RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Modary C., Morris J., Mostrelli A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson K.A., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of *Drosophila melanogaster*,"  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003698; AAF54888.2; -  
 DR Flybase: FBgn0038108; CG7518.  
 DR InterPro: IPR001005; Myb\_DNA\_bind.  
 DR PROSITE: PS00037; MYB\_1; UNKNOWN\_1.  
 SQ SEQUENCE 2451 AA; 266959 MW; 086A2293F27481E2 CRC64;

Query Match 77.1%; Score 84; DB 5; Length 2451;  
 Best Local Similarity 91.3%; Pred. No. 1.5;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 AAAAAAAAAAAAAAAAAAAAAA 25  
 DB 1374 AAAAAAAAAAAAAAAAAAAAAA 1396

RESULT 2  
 Q9EPW8 PRELIMINARY: PRT; 1354 AA.  
 AC Q9EPW8;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE NISCHARIN.  
 GN NISCH.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ICR OUTBRED; TISSUE=BRIN;  
 RX MEDLINE=20571837; PubMed=11121431;  
 RA Alahari S.K., Lee J.W., Juliano R.L.;  
 RT "Nischarin, a Novel Protein That Interacts with the Integrin  $\alpha$ 5 $\beta$   
 RT Subunit and Inhibits Cell Migration,"  
 RL J. Cell Biol. 151:1141-1154(2000).  
 DR EMBL: AF315344; AAG42100.1; -  
 DR MGP: MGI:1928323; Nisch.  
 DR InterPro: IPR001128; Cyt\_P450.  
 DR InterPro: IPR001611; LRR.  
 DR InterPro: IPR003592; LRR\_out.  
 DR Pfam: PF00560; LRR\_4.  
 DR PRINTS: PR00019; LEURICRPT.  
 DR SMART: SMO0370; LRR\_5.  
 DR PROSITE: PS00086; CYTOCHROME\_P450; UNKNOWN\_1.  
 SQ SEQUENCE 1354 AA; 148060 MW; 01BD675FDCAI9247 CRC64;

Query Match 73.4%; Score 80; DB 11; Length 1354;  
 Best Local Similarity 83.3%; Pred. No. 2.2;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAAAAAA 25  
 DB 837 EAPAAAEAPAAAEAPAAAEAPAA 860

RESULT 3  
 Q91WU0 PRELIMINARY: PRT; 110 AA.  
 AC Q91WU0;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL 10.7 KDA PROTEIN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Salivary Gland;  
 RA Strausberg R.;  
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC012681; AAH12681.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 110 AA; 10662 MW; 0581D2F635F87EAB CRC64;

Query Match 72.5%; Score 79; DB 11; Length 110;  
 Best Local Similarity 83.3%; Pred. No. 0.32;  
 Matches 20; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAAAAAA 25  
 DB 23 DAAAAAAAAAAAAAAAAAAAAA 46

RESULT 4  
 Q39598 PRELIMINARY: PRT; 265 AA.  
 AC Q39598;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE CGCR-4. PRODUCT (FRAGMENT).  
 GN CGCR-4.  
 OS Chlamydomonas reinhardtii.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Chlamydomonadaceae; Chlamydomonas.  
 OX NCBI\_TaxID=3055;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92119224; PubMed=1731966;  
 RA Wakarchuk W.W., Muller F.W., Beck C.F.;  
 RT "Two GC-rich DNA elements of *Chlamydomonas reinhardtii* with complex  
 RT arrangements of directly repeated sequence motifs,"  
 RL Plant Mol. Biol. 18:143-146(1992).  
 DR EMBL: X17208; CAA35080.1; -  
 FT NON\_TER  
 SQ SEQUENCE 265 AA; 26216 MW; B35318B7377CF782 CRC64;

Query Match 70.6%; Score 77; DB 10; Length 265;  
 Best Local Similarity 82.6%; Pred. No. 1;  
 Matches 19; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 3 AAAAAAAAAAAAAAAAAAAAAA 25  
 DB 154 AAAAAAAAAAKARVAAEAPAA 176

RESULT 5  
 Q9QXG2 PRELIMINARY: PRT; 665 AA.  
 AC Q9QXG2;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)



DT 01-JUN-2001 (Tremblrel. 17, last annotation update)  
DE CHORODEREMIA PROTEIN.  
GN CEM.  
OS Mus musculus (mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C;  
RA van den Hurk J.A., Huber I., van de Pol T.J., Cremers F.P.;  
RT Cloning and sequencing of the mouse choroaderemia gene."  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF218084; AAF25478.1; -  
DR MGD; MGI:892979; Chm.  
DR InterPro: IPR002005; Rab\_GDI\_REP.  
DR Pfam: PF00996; GDI; 1.  
DR PRINTS: PRO0891; RABGDIREP.  
SQ SEQUENCE 665 AA; 73976 MW; FF71A74AD3FBDE0A CRC64;

Query Match 69.7%; Score 76; DB 11; Length 665;  
Best Local Similarity 79.2%; Pred. No. 2.8;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAAERAAAAERAAAAA 25  
DB 132 EAARAAERAAERAAERAAEA 155

RESULT 6  
OQ9H4A0 PRELIMINARY; PRT; 1452 AA.  
AC OQ9H4A0;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)  
DE CDC2L5 PROTEIN KINASE.  
GN CDC2L5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RA Marques F., Moreau J.L., Peaucellier G., Lozano J.C., Schatt P.,  
RT Picard A., Callebaut I., Perre E., Genevriere A.M.;  
RT "A new subfamily of high molecular mass CDC2-related kinases with  
RT PITAI/VRP."  
RL Biochem. Biophys. Res. Commun. 279:832-837(2001).  
DR EMBL; AJ297710; CAC10401.1; -  
DR HSSP; P24941; 1BUH.  
DR InterPro: IPR000719; Euk\_PKinase.  
DR InterPro: IPR002290; Ser\_Thr\_PKinase.  
DR InterPro: IPR001245; Tyr\_PKinase.  
DR Pfam: PF00069; Pkinase; 1.  
DR SMART; SM00220; S\_TKc; 1.  
DR SMART; SM00219; Tyrc; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW ATP-binding; Kinase; Transferase.  
SQ SEQUENCE 1452 AA; 158480 MW; C7ED072968B439CB CRC64;

Query Match 69.7%; Score 76; DB 4; Length 1452;  
Best Local Similarity 79.2%; Pred. No. 5.5;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAAERAAAAERAAAAA 25  
DB 463 EAARAAERAAERAAERAAEA 486

RESULT 7  
OQ9H4A1 PRELIMINARY; PRT; 1512 AA.  
AC OQ9H4A1;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)  
DE CDC2L5 PROTEIN KINASE.  
GN CDC2L5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RA Marques F., Moreau J.L., Peaucellier G., Lozano J.C., Schatt P.,  
RT Picard A., Callebaut I., Perre E., Genevriere A.M.;  
RT "A new subfamily of high molecular mass CDC2-related kinases with  
RT PITAI/VRP."  
RL Biochem. Biophys. Res. Commun. 279:832-837(2001).  
DR EMBL; AJ297709; CAC10400.1; -  
DR HSSP; P24941; 1BUH.  
DR InterPro: IPR000719; Euk\_PKinase.  
DR InterPro: IPR002290; Ser\_Thr\_PKinase.  
DR InterPro: IPR001245; Tyr\_PKinase.  
DR Pfam: PF00069; Pkinase; 1.  
DR SMART; SM00220; S\_TKc; 1.  
DR SMART; SM00219; Tyrc; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW ATP-binding; Kinase; Transferase.  
SQ SEQUENCE 1512 AA; 164969 MW; 283B80553DB57650 CRC64;

Query Match 69.7%; Score 76; DB 4; Length 1512;  
Best Local Similarity 79.2%; Pred. No. 5.7;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAAERAAAAERAAAAA 25  
DB 463 EAARAAERAAERAAERAAEA 486

RESULT 8  
OQ4426 PRELIMINARY; PRT; 531 AA.  
AC OQ4426;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-MAY-1999 (Tremblrel. 10, last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)  
DE TROPOMYOSIN ISOFORM 33.  
GN TMI OR G4898.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=84205681; PubMed=6202423;  
RA Karik C.C., Mahaffey J.W., Coutu M.D., Fyberg E.A.;  
RT "Organization of contractile protein genes within the 88F subdivision  
RT of the D. melanogaster third chromosome."  
RL Cell 37:469-481(1984).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87064486; PubMed=3097506;  
RA Karik C.C., Fyberg E.A.;  
RT "Two Drosophila melanogaster tropomyosin genes: structural and  
RT functional aspects."  
RL Mol. Cell. Biol. 6:1965-1973(1986).



RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Rehner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spletter E., Spindling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003503; AAF4641.1;  
 DR FlyBase: FBgn030769; CG13012.  
 SQ SEQUENCE 221 AA; 22987 MW; A1B959198167C5E2 CRC64;

Query Match 68.8%; Score 75; DB 5; Length 221;  
 Best Local Similarity 81.8%; Pred. No. 1.4;  
 Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 4 AAAAAAAAAAAAAAAAAA 25  
 Db 6 AAAAAAAAAAAATVAEAAVA 27

RESULT 11  
 ID Q9W2J2 PRELIMINARY; PRT: 1071 AA.  
 AC Q9W2J2; 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 GN CG18375. PROTEIN.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RA MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Morten J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Flannkuch C., Baldwin D.,  
 RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Bortman M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Chertys J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.K.,  
 RA Dodson K., Doup L.E., Domes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foslter C., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegyan C.,  
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Rehner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spletter E., Spindling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003453; AAF46599.1;  
 DR HSSP: Q13625; IYCS.  
 DR FlyBase: FBgn0034606; CG18375.  
 DR InterPro: IPR002110; ANK.  
 DR InterPro: IPR001452; SH3.  
 DR Pfam: PF00023; ank; 2.  
 DR Pfam: PF00018; SH3; 1.  
 DR SMART: SM00248; ANK; 2.  
 DR SMART: SM00326; SH3; 1.  
 DR PROSITE: PS50088; ANK\_REPEAT; 2.  
 DR PROSITE: PS50297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE: PS50002; SH3; 1.  
 KW ANK repeat, Repeat.  
 SQ SEQUENCE 1071 AA; 115502 MW; F86840E92692B4FA CRC64;

Query Match 66.1%; Score 72; DB 5; Length 1071;  
 Best Local Similarity 79.2%; Pred. No. 10;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAA 25  
 Db 504 EAAAAAAAAAAAAAAAAAQAEEAA 527

RESULT 12  
 ID Q39597 PRELIMINARY; PRT: 206 AA.  
 AC Q39597; 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 GN CGCR-1. PRODUCT (FRAGMENT).  
 OS CGCR-1.  
 OC Chlamydomonas reinhardtii.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Chlamydomonadaceae; Chlamydomonas.  
 OX NCBI\_TaxID=3055;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CM15 MT-;  
 RA MEDLINE=92119224; PubMed=1731966;  
 RA Wakarchuk W.W., Muller F.W., Beck C.F.,  
 RT "Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex  
 RT arrangements of directly repeated sequence motifs.";  
 RL Plant Mol. Biol. 18:143-146(1992).  
 DR EMBL: X17207; CA835079.1;  
 DR InterPro: IPR001778; POA\_allergen.  
 DR PRINTS: PR00833; POAALLERGEN.  
 FT NON\_TER 1  
 FT TER 206  
 SQ SEQUENCE 206 AA; 19869 MW; ED3FF120EF8EFAE1 CRC64;

Query Match 65.1%; Score 71; DB 10; Length 206;  
 Best Local Similarity 75.0%; Pred. No. 3.1;  
 Matches 18; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAA 25  
 Db 49 EAAAAAAAAAAAAAAAAAERAPPA 72

```

RESULT 13
Q9SWH3 PRELIMINARY; PRT: 1216 AA.
AC Q9SWH3;
DT 01-MAY-2000 (TRMBLrel. 13, Created)
DT 01-MAY-2000 (TRMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TRMBLrel. 19, Last annotation update)
DE VARIABLE FLAGELLAR NUMBER PROTEIN.
GN VFLI.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
CC Chlamydomonadaceae; Chlamydomonas.
CX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-21GR, CC-1690;
RA MEDLINE=94063478; PubMed=8244002;
RT Tam L.W., Lefebvre P.A.;
RT "Cloning of flagellar genes in Chlamydomonas reinhardtii by DNA
RT insertional mutagenesis."
RL Genetics 135:375-384(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-21GR, CC-1690;
RA MEDLINE=21181843; PubMed=11285274;
RT Sifflow C.D., Lavoie M., Tam L.W., Tousey S., Sanders M., Wu W.C.,
RT Borodovsky M., Lefebvre P.A.;
RT "The vfl protein in Chlamydomonas localizes in a rotationally
RT asymmetric pattern at the distal ends of the basal bodies."
RL J. Cell Biol. 153:63-74(2001).
DR EMBL: AF154916; AAD52203.1; -
DR InterPro: IPR004089; Chemotaxis_transducer.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR_out.
DR Pfam: PF005560; LRR.
DR SMART: SM00370; LRR.
KW Flagella.
SQ SEQUENCE 1216 AA; 127943 MW; F06E798B35AF256E CRC64;

Query Match 65.1%; Score 71; DB 10; Length 1216;
Best Local Similarity 73.9%; Pred. No. 14;
Matches 17; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 EAAAAAEEAAAAEAAAAA 24
Db 635 EAAAGREMAAAYVOAAAAELAVA 657

RESULT 14
Q9M4X9 PRELIMINARY; PRT: 1787 AA.
AC Q9M4X9;
DT 01-OCT-2000 (TRMBLrel. 15, Created)
DT 01-OCT-2000 (TRMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TRMBLrel. 19, Last annotation update)
DE FLAGELLAR AUTOTOMY PROTEIN FAIP.
GN FAI.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
CC Chlamydomonadaceae; Chlamydomonas.
CX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RA First R.V., Kim P.J., Griffiths E.R., Quarby L.M.;
RT "Faip is a 171 kDa protein essential for axonemal microtubule severing
RT in Chlamydomonas."
RL J. Cell Sci. 0:0-0(2000).
DR EMBL: AF246990; AAF66419.1; -
DR InterPro: IPR003592; LRR_out.
DR SMART: SM00370; LRR.
KW
```

```

SQ SEQUENCE 1787 AA; 171584 MW; 452A74259E14DC1 CRC64;

Query Match 65.1%; Score 71; DB 10; Length 1787;
Best Local Similarity 81.8%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAAAAEEAAAAEAAAAA 24
Db 1704 AAAAMEAARAMEAAAALEAAAA 1725

RESULT 15
Q9H782 PRELIMINARY; PRT: 323 AA.
AC Q9H782;
DT 01-MAR-2001 (TRMBLrel. 16, Created)
DT 01-MAR-2001 (TRMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TRMBLrel. 19, Last annotation update)
DE CDNA: FLJ21157 FIS, CLONE CAS09937 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Ohbayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isoigai T., Sugano S.;
RT "NEO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AK024810; BAB15016.1; -
DR InterPro: IPR000104; Antifreeze_1.
DR InterPro: IPR002713; FF.
DR InterPro: IPR001202; WW.
DR Pfam: PF01846; FF.
DR PRINTS: PR00308; ANTIFREEZE1.
DR SMART: SM00441; FF.
DR SMART: SM00456; WW.
DR PROSITE: PS01159; WW_DOMAIN_1; 1.
DR PROSITE: PS50020; WW_DOMAIN_2; 2.
FT NON_TER 323
FT SEQUENCE 323 AA; 35148 MW; EAC7B36489FA074A CRC64;

Query Match 64.2%; Score 70; DB 4; Length 323;
Best Local Similarity 85.7%; Pred. No. 5.6;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 AAAAAAEEAAAAEAAAAA 25
Db 141 AAAEAAAAAYVAAAAAALAAAA 161
```

Search completed: July 1, 2002, 06:30:57  
Job time: 689 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:19:22 ; Search time 98.97 Seconds

(without alignments)  
28.057 Million cell updates/sec

Title: US-09-461-684-2

Perfect score: 109  
Sequence: 1 CEAAAAEAAAAEAAAAEAAAAA 25

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

```

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	109	100.0	25	21	AA13781
2	109	100.0	45	21	AA13783
3	84	77.1	2451	22	ABR71574
4	79	72.5	119	22	ABBA0362
5	79	72.5	119	22	AAW61172
6	79	72.5	119	22	AAW73866
7	79	72.5	119	22	AAW34068
8	76	69.7	112	22	ABG15501
9	76	69.7	737	11	AAK05268
10	75.5	69.3	566	22	ABR61040
11	75	68.8	221	22	ABR67690

12	73	67.0	225	21	AAK00758	Human secreted pro
13	72	66.1	54	18	AAW22875	P. americanus skin
14	72	66.1	316	22	ABG18917	Novel human diagno
15	72	66.1	1071	22	ABR60579	Drosophila melanog
16	72	66.1	1444	22	ABG15667	Novel human diagno
17	70	64.2	39	18	AAW22874	P. americanus skin
18	70	64.2	326	22	AAW60765	Gene 16 related pe
19	70	64.2	326	22	AAW60766	Human polypeptide
20	70	64.2	487	22	AAW41402	Human WT1 interact
21	70	64.2	957	22	AAW54150	Synthetic skin typ
22	69	63.3	39	18	AAW22876	P. americanus anti
23	69	63.3	39	18	AAW22881	Novel human diagno
24	69	63.3	262	22	ABG15586	CORR potassium cha
25	69	63.3	1529	17	AAW97985	Drosophila melanog
26	69	63.3	1669	22	ABR54003	P. americanus skin
27	68	62.4	39	18	AAW22871	P. americanus skin
28	68	62.4	39	18	AAW22869	Drosophila melanog
29	68	62.4	148	22	ABR70649	Novel human diagno
30	68	62.4	1078	22	ABR58620	Drosophila melanog
31	68	62.4	2023	22	ABR53487	Human neuroblastom
32	67	61.5	314	18	AAW14283	Mouse Phox2b prote
33	67	61.5	314	21	AAW70572	Drosophila melanog
34	67	61.5	544	22	ABR51602	Drosophila melanog
35	67	61.5	1037	22	ABR70288	Drosophila melanog
36	67	61.5	1416	22	ABR55499	Drosophila melanog
37	67	61.5	2703	22	ABR53299	Peptide modulating
38	67	61.5	31	21	AAW8166	Elmeria gametocyte
39	66	60.6	76	17	AAW91273	Amino acid fragmen
40	66	60.6	76	20	AAW77190	Shorthorn sculpin
41	66	60.6	92	21	AAW44712	Human protein sequ
42	66	60.6	372	22	ABR94062	Drosophila melanog
43	66	60.6	399	22	ABR37773	Drosophila melanog
44	66	60.6	634	22	ABR71624	Drosophila melanog
45	66	60.6	924	22	ABR67870	Drosophila melanog

#### ALIGNMENTS

RESULT	ID	Description
1	AA13781	
AA13781	AA13781	standard; peptide: 25 AA.
10-NOV-2000	AA13781	(First entry)
Soluble peptide antigen pEA.		
peptide: cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy; major histocompatibility complex class I; MHC class I; antigen; tumour; prostate; breast; multiple myeloma.		
Unidentified.		
WO200035949-A1.		
22-JUN-2000.		
14-DEC-1999;	99WO-US29724.	
14-DEC-1998;	98US-0112324.	
(DEND-) DENDREON CORP.		
Laus R, Hakim I, Vidovic D;		
WPI: 2000-442365/38.		
Antigens modified by the covalent addition of a peptide that facilitates entry into antigen presenting cells, useful for producing compositions for immunizing against tumors and pathogens		

PS Claim 2; Page 26; 34pp: English.

CC The present invention relates to compositions of modified soluble protein  
CC antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)  
CC response i.e. a major histocompatibility complex (MHC) class I molecule  
CC peptide sequence which facilitate entry of the antigen into antigen  
CC presenting cells (APCs). The present sequence is one such peptide  
CC sequence which can be used to modify the soluble antigens. The present  
CC sequence is peptide pEA. The modified antigen composition may be used for  
CC immunising against, or treating a tumour e.g. prostate and breast  
CC carcinoma or multiple myeloma, or pathogen in mammals.

XX Sequence 25 AA;

Query Match 100.0%; Score 109; DB 21; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1.8e-07;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CEAFAAEEAFAAFAAFAAFAA 25  
1 ccaaaaaaaaaaaaaaaaaaaaaa 25

RESULT 2  
AAB13783  
ID AAB13783 standard; peptide; 45 AA.

XX AAB13783;

DT 10-NOV-2000 (first entry)

DE Soluble tandem pEA/ PK peptide conjugate.

XX PK peptide; cytotoxic; vaccine; cytotoxic T cell; CTL; immunotherapy;  
KM major histocompatibility complex class I; MHC class I; antigen; tumour;  
KW prostate; breast; multiple myeloma; pEA peptide.

XX Unidentified.

OS WO200035949-A1.

PN 22-JUN-2000.

PF 14-DEC-1999; 99WO-US29724.

PR 14-DEC-1998; 98US-0112324.

XX (DEND-) DENDREON CORP.

PI Laus R, Hakim I, Vidovic D;

XX WPI; 2000-442365/38.

PT Antigens modified by the covalent addition of a peptide that

PT facilitates entry into antigen presenting cells, useful for producing

PT compositions for immunizing against tumors and pathogens -

PS Claim 2; Page 26; 34pp: English.

XX The present invention relates to compositions of modified soluble protein  
CC antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)  
CC response i.e. a major histocompatibility complex (MHC) class I molecule  
CC peptide sequence which facilitate entry of the antigen into antigen  
CC presenting cells (APCs). The present sequence is one such peptide  
CC sequence which can be used to modify the soluble antigens. The present  
CC sequence is tandem pEA/ PK peptide conjugate. The modified antigen  
CC composition may be used for immunising against, or treating a tumour e.g.  
CC prostate and breast carcinoma or multiple myeloma, or pathogen in  
XX mammals.

SQ Sequence 45 AA;

Query Match 100.0%; Score 109; DB 21; Length 45;  
Best Local Similarity 100.0%; Pred. No. 3.2e-07;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CEAFAAEEAFAAFAAFAAFAA 25  
1 ccaaaaaaaaaaaaaaaaaaaaaa 25

RESULT 3  
ABB71574  
ID ABB71574 standard; Protein; 2451 AA.

XX ABB71574;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 41514.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL15677.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

PS Disclosure; SEQ ID NO 41514; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABB5737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2451 AA;

Query Match 77.1%; Score 84; DB 22; Length 2451;

Best Local Similarity 91.3%; Pred. No. 0.021; 2; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 AAAAEEAFAAFAAFAAFAAFAA 25  
1374 aaaaaaaaaaaaaaaaaaaaaa 1396

```
RESULT 4
ABB40362
ID ABB40362 standard; Peptide; 119 AA.
XX
XX ABB40362;
AC
XX
XX 04-FEB-2002 (first entry)
DT
XX
DE Peptide #7868 encoded by human foetal liver single exon probe.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
XX Homo sapiens.
OS
XX WO200157275-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00669.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-483447/52.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
XX Claim 27; SEQ ID NO 32997; 639pp + sequence listing; English.
PS
XX
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC
XX
XX Sequence 119 AA;
SQ
```

```
Query Match 72.5%; Score 79; DB 22; Length 119;
Best Local Similarity 83.3%; Pred. No. 0.0043;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
CY 2 EAAAAEAAAAEAAAAEAAAAA 25
||| | ||||| ||||| |||||
DB 31 eaeaeaeaeaeaeaeaeaeae 54
```

```
RESULT 5
AAM61172
ID AAM61172 standard; Protein; 119 AA.
XX
XX AAM61172;
AC
XX
XX 05-NOV-2001 (first entry)
DT
XX
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 33277.
DE
XX
XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW
```

```
KN epilepsy; cancer.
XX
XX Homo sapiens.
OS
XX WO200157275-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00667.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-483446/52.
DR
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
XX Example 4; SEQ ID NO: 33277; 650pp + Sequence Listing; English.
PS
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
CC
XX
XX Sequence 119 AA;
SQ
```

```
Query Match 72.5%; Score 79; DB 22; Length 119;
Best Local Similarity 83.3%; Pred. No. 0.0043;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
CY 2 EAAAAEAAAAEAAAAEAAAAA 25
||| | ||||| ||||| |||||
DB 31 eaeaeaeaeaeaeaeaeaeae 54
```

```
RESULT 6
AAM73886
ID AAM73886 standard; Protein; 119 AA.
XX
XX AAM73886;
AC
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 34192.
DE
XX
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
XX Homo sapiens.
OS
XX WO200157276-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00668.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
```



PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234657.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488900/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human bone marrow -  
 XX  
 PS Example 4; SEQ ID NO: 34192; 658bp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukemia and myeloma. The present sequence is a  
 CC protein encoded by one of the probes of the invention.  
 CC  
 XX  
 SQ Sequence 119 AA;

Query Match 72.5%; Score 79; DB 22; Length 119;  
 Best Local Similarity 83.3%; Pred. No. 0.0043;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 EAAAAEAAAAEAAAAEAAAAA 25  
 || | ||||| ||||| |||||  
 DB 31 eaeaeaeaeaeaeaeaeaeaeae 54

RESULT 7  
 AAM34068  
 ID AAM34068 standard; Protein: 119 AA.  
 XX  
 AC AAM34068;  
 XX  
 DT 17-OCT-2001 (first entry)  
 XX  
 DE Peptide #9105 encoded by probe for measuring placental gene expression.  
 XX  
 KW Probe; microarray; human; placenta; antenatal diagnosis;  
 KW genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157272-A2.  
 PN  
 PD 09-AUG-2001.  
 PD  
 PE 30-JAN-2001; 2001WO-US00663.  
 PE  
 XX 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234657.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 PR  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488900/53.  
 DR  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human placenta -  
 XX  
 CC

PS Claim 27; SEQ ID NO 34337; 654bp; English.  
 XX  
 CC The present invention relates to single exon nucleic acid probes (SENP:  
 CC see AAI31315-AA157546). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for producing a microarray for  
 CC predicting, measuring and displaying gene expression in samples derived  
 CC from human placenta. The probes are useful for antenatal diagnosis of  
 CC human genetic disorders.  
 CC  
 XX  
 SQ Sequence 119 AA;  
 XX  
 Query Match 72.5%; Score 79; DB 22; Length 119;  
 Best Local Similarity 83.3%; Pred. No. 0.0043;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 EAAAAEAAAAEAAAAEAAAAA 25  
 || | ||||| ||||| |||||  
 DB 31 eaeaeaeaeaeaeaeaeaeaeae 54

RESULT 8  
 ABG15501  
 ID ABG15501 standard; Protein: 112 AA.  
 XX  
 AC ABG15501;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #15492.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 PN  
 PD 11-OCT-2001.  
 PD  
 PE 30-MAR-2001; 2001WO-US08631.  
 PE  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 PR  
 PA (HYSE-) HYSBO INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YR;  
 XX  
 DR WPI; 2001-639362/73.  
 DR  
 DR N-PSDB; AAS79688.  
 DR  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PT  
 XX Claim 20; SEQ ID NO 45860; 103bp; English.  
 XX  
 PS The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 112 AA;  
 OY 3 AAAAAAAAAAAAAAAAAAAAAA 25  
 DB 82 aaaaaaaaaaaaaaaaaaaaaa 104  
 RESULT 9  
 ID AAR05268 standard: protein; 737 AA.  
 AC AAR05268;  
 XX  
 XX 15-AUG-1990 (first entry)  
 DE Amino acid sequence of human megakaryocytopoietin (MKP)  
 DE encoded by upper reading frame of DNA contained in clone 14.  
 XX  
 KW Human megakaryocytopoietin (MPK); haematopoiesis; immunoassay;  
 KW rare leukaemia diagnosis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP354989-A.  
 PN JF03195496-A.  
 XX  
 PD 21-FEB-1990.  
 XX  
 PF 27-JUN-1989; 89UP-0111714.  
 XX  
 PR 28-JUN-1988; 88US-0212623.  
 XX  
 PA (YISS ) YISSUM RES DEV CO.  
 XX  
 PI Hermona S, Haim Z;  
 XX  
 XX WPI: 1990-052749/08.  
 DR WPI: 1991-282630/40.  
 DR N-PDB: AAQ92263.  
 XX  
 XX New polypeptide hormone megakaryocytopoietin -  
 PT involved in haematopoiesis and immune response, and derived  
 PT nucleic acid, antibodies, etc., useful eg in diagnosis of leukaemia  
 XX  
 PS Claim 7, Page 17; Fig 1; 24pp; English.  
 XX  
 CC A CDNA library constructed from foetal ganglioside mRNA was screened  
 CC with a butyryl-cholesterase (bch) cDNA probe to identify one clone  
 CC (clone 14) contg. about 250 bases at the 5'-region of the bch coding  
 CC sequence plus a sequence which hybridised with genomic DNA from patients  
 CC with acute myelocytic leukaemia. AAP95268 is the translation of the  
 CC upper reading frame and is claimed in the patent. Its nucleic acid can  
 CC be isolated, opt. together with its signal sequence, by screening cDNA  
 CC or genomic libraries with the clone 14 DNA, to identify a full-length  
 CC clone. The material from this clone can be transferred into mammalian or  
 CC microbial host cells and these cultured for its prodn. It has hormonal  
 CC activity in modulation of haematopoiesis and immune responses. It is  
 CC useful in standard immunoassay or hybridisation procedures for  
 CC classification and diagnosis of rare leukaemias. It may also be useful  
 CC therapeutically.

XX  
 SQ Sequence 737 AA;  
 OY 2 EAAAAAAAAAAAAAAAAAAAAA 25  
 DB 106 eaataaakaatkaataaaka 129  
 RESULT 10  
 ID ABB61040 standard: Protein; 566 AA.  
 AC ABB61040;  
 XX  
 XX 26-MAR-2002 (first entry)  
 DE Drosophila melanogaster polypeptide SEQ ID NO 9912.  
 DE Drosophila melanogaster polypeptide SEQ ID NO 9912.  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 PN WO2001/1042-A2.  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PERE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 XX WPI: 2001-656860/75.  
 DR N-PDB: ABL05143.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Disclosure; SEQ ID NO 9912; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB1616175) and the encoded proteins  
 CC (ABBS7737-ABR12072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 566 AA;  
 OY 2 EAAAAAAAAAAAAAAAAAAAAA 25  
 DB 362 eaataaadaae-aadaaadaa 384  
 Query Match 69.3%; Score 75.5; DB 22; Length 566;  
 Best Local Similarity 87.5%; Pred. No. 0.054;  
 Matches 21; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```
RESULT 11
XX ABB67690
ID ABB67690 standard; Protein; 221 AA.
XX
XX ABB67690;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 29862.
DE
XX
XX Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PT
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX
XX (PEKE ) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EM.
PI
XX
XX WPI: 2001-656860/75.
DR
XX
XX N-PSDB; ABL11793.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PR interactions -
XX
XX
XX Disclosure: SEQ ID NO 29862; 21bp + Sequence Listing: English.
XX
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB120511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins
CC (AB157737-AB172072).
CC
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcr_sequences.
CC
XX
XX Sequence 221 AA;
SQ
Query Match 68.8%; Score 75; DB 22; Length 221;
Best Local Similarity 81.8%; Pred. NO. 0.025;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 4 AAAAEEAAAAEAAAAEAAAA 25
DB 6 aaaaaaavaaataaavaa 27
RESULT 12
XX AAG00758
ID AAG00758 standard; Protein; 225 AA.
XX
XX AAG00758;
AC
XX
XX 06-OCT-2000 (first entry)
DT
XX
XX Human secreted protein, SEQ ID NO: 4839.
DE
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW
```

```
KW gene therapy; chromosome mapping.
XX
XX Homo sapiens.
OS
XX
XX EP1033401-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 21-FEB-2000; 2000EP-0200610.
PE
XX
XX 26-FEB-1999; 99US-0122487.
PR
XX
XX (GEST ) GENSET.
PA
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI
XX
XX WPI: 2000-500381/45.
DR
XX
XX N-PSDB; AAC00764.
XX
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PR diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 13; SEQ ID 4839; 71bp + CD-ROM; English.
XX
XX
XX The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
CC
XX
XX Sequence 225 AA;
SQ
Query Match 67.0%; Score 73; DB 21; Length 225;
Best Local Similarity 86.4%; Pred. NO. 0.045;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 3 AAAAEEAAAAEAAAAEAAAA 24
DB 150 aaaaaaaaaaaaaaaaaa 171
RESULT 13
XX AAW22875
ID AAW22875 standard; Protein; 54 AA.
XX
XX AAW22875;
AC
XX
XX 23-MAR-1998 (first entry)
DT
XX
XX P. americanus skin type antifreeze polypeptide, sAFP8.
DE
XX
XX Antifreeze polypeptide; intracellular skin type AFP; winter flounder;
KW cold resistance; thermal hysteresis; antibacterial.
XX
XX Pseudopleuronectes americanus.
OS
XX
XX WO9728260-A1.
PN
XX
XX 07-AUG-1997.
PD
XX
XX 30-JAN-1997; 97WO-CA00062.
PF
XX
XX 31-JAN-1996; 96US-0010920.
PR
```

XX (HSCR-) HSC RES & DEV LP.  
 PA  
 XX Gong Z, Hew C;  
 PI  
 XX WPI: 1997-402614/37.  
 DR N-PSDB; AAT75505.  
 XX  
 PT Isolated fish skin anti-freeze polypeptide - useful to depress  
 PT freezing points of aqueous compositions and protect plant, fungal,  
 PT animal or bacterial cells from cold  
 PS  
 XX Claim 6; Page 68; 104pp; English.  
 XX  
 CC This sequence represents the sAPP8 clone corresponding to a novel skin  
 CC intracellular antifreeze polypeptide (AFP) which induces a concentration  
 CC dependent decrease in the freezing point of an aqueous solution. This  
 CC novel polypeptide does not contain a signal sequence and is thought to be  
 CC intracellular. AFP's can be used to make an aqueous composition resistant  
 CC to freezing by changing its thermal hysteresis such as a water or salt  
 CC solution, an intracellular compartment of a cell or a food stuff, e.g.  
 CC soft serve 'frozen' yoghurt or ice cream. AFP's can inhibit ice  
 CC recrystallisation during cold storage, improving the texture and  
 CC palatability of the food and has antibacterial properties. Such  
 CC polypeptides can also be expressed to provide cold resistance to cells,  
 CC e.g. plant, fungal animal or bacterial cells. The antibodies can be used  
 CC to identify and isolate AFP while its promoter can be used to direct  
 CC expression of a nucleic acid.  
 XX  
 SO Sequence 54 AA;  
 XX  
 Query Match 66.1%; Score 72; DB 18; Length 54;  
 Best Local Similarity 78.3%; Pred. No. 0.014;  
 Matches 18; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 3 AAAAAEAAAEEAAAEEAAA 25  
 DB 28 aaaaataaaaakaaaakaaaa 50  
 XX  
 RESULT 14  
 ABG18917  
 ID ABG18917 standard; Protein; 316 AA.  
 XX  
 AC ABG18917;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #18908.  
 XX  
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI: 2001-639362/73.  
 DR N-PSDB; AAS83104.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX  
 XX Claim 20; SEQ ID NO 49276; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SO Sequence 316 AA;  
 XX  
 Query Match 66.1%; Score 72; DB 22; Length 316;  
 Best Local Similarity 82.6%; Pred. No. 0.083;  
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 3 AAAAAEAAAEEAAAEEAAA 25  
 DB 14 aaaaavaaaaaaavaaaaa 36  
 XX  
 RESULT 15  
 ABB60579  
 ID ABB60579 standard; Protein; 1071 AA.  
 XX  
 AC ABB60579;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 8529.  
 XX  
 KM Drosophila; developmental biology; cell signalling; insecticide;  
 KM pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li FWD, Myers EW;  
 XX  
 DR WPI: 2001-656860/75.  
 DR N-PSDB; ABL04682.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT Interactions -

XX PS Disclosure; SEQ ID NO 8529; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent  
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX

XX Sequence 1071 AA;

Query Match 66.1%; Score 72; DB 22; Length 1071;

Best Local Similarity 79.2%; Pred. NO. 0.28; Mismatches 5; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 EAAAAAERAAAAERAAAAERAAAA 25

Db 504 eaaaaaaaaaaaaaaaaaaaaaa 527

Search completed: July 1, 2002, 06:19:23  
Job time: 510 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:28:10 ; Search time 49.45 Seconds  
(without alignments)  
46.636 Million cell updates/sec

Title: US-09-461-684-3

Perfect score: 143

Sequence: 1 GGFCAIAGFIENGEGMIDGMYG 24

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 segs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	134	93.7	550	1 HMIIVS2	hemagglutinin prec
2	134	93.7	550	1 HMIIVS3	hemagglutinin prec
3	134	93.7	550	1 HMIIV77	hemagglutinin prec
4	134	93.7	550	1 HMIIV80	hemagglutinin prec
5	134	93.7	550	1 HMIIV33	hemagglutinin prec
6	134	93.7	550	1 HMIIV89	hemagglutinin prec
7	134	93.7	550	1 HMIIV21	hemagglutinin prec
8	134	93.7	550	1 HMIIV98	hemagglutinin prec
9	134	93.7	550	1 HMIIV15	hemagglutinin prec
10	134	93.7	550	2 JQ1153	hemagglutinin prec
11	134	93.7	550	2 JQ1154	hemagglutinin prec
12	134	93.7	550	2 JQ1155	hemagglutinin prec
13	134	93.7	556	1 HMTVH	hemagglutinin prec
14	134	93.7	556	1 HMTVHA	hemagglutinin prec
15	134	93.7	556	1 HMTVHM	hemagglutinin prec
16	134	93.7	556	1 HMTVDU	hemagglutinin prec
17	133	93.0	561	1 HMTIV49	hemagglutinin prec
18	133	93.0	561	1 HMTIV84	hemagglutinin prec
19	132	92.3	565	1 HMTIVEL	hemagglutinin prec
20	132	92.3	565	1 HMTIVE3	hemagglutinin prec
21	132	92.3	566	1 HMTIV6	hemagglutinin prec
22	132	92.3	567	1 HMTIV	hemagglutinin prec
23	131	91.6	362	2 S38637	hemagglutinin - in
24	131	91.6	550	1 HMTIV86	hemagglutinin prec
25	131	91.6	550	1 HMTIVT7	hemagglutinin prec
26	131	91.6	565	1 HMTIVE2	hemagglutinin prec
27	131	91.6	565	1 HMTIVE4	hemagglutinin prec
28	131	91.6	565	1 HMTIVE5	hemagglutinin prec
29	131	91.6	565	1 HMTIVE6	hemagglutinin prec

30	131	91.6	565	1 HMTIVE7	hemagglutinin prec
31	131	91.6	565	1 HMTIVE8	hemagglutinin prec
32	131	91.6	565	1 HMTIVE9	hemagglutinin prec
33	131	91.6	565	1 HMTIVT	hemagglutinin prec
34	131	91.6	565	1 HMTIVE	hemagglutinin prec
35	131	91.6	565	2 S33703	hemagglutinin - in
36	131	91.6	570	1 A45591	hemagglutinin prec
37	131	91.6	570	2 S22013	hemagglutinin prec
38	131	91.6	570	2 S22014	hemagglutinin prec
39	131	91.6	570	2 S22015	hemagglutinin prec
40	131	91.6	570	2 S22016	hemagglutinin prec
41	131	91.6	570	2 S22017	hemagglutinin prec
42	131	91.6	570	2 S22018	hemagglutinin prec
43	131	91.6	570	2 S22020	hemagglutinin prec
44	131	91.6	570	2 S22021	hemagglutinin prec
45	131	91.6	570	2 S22029	hemagglutinin prec

#### ALIGNMENTS

RESULT 1  
HMTIVS2  
hemagglutinin precursor - influenza A virus (strain A/swine/126/82) (fragment)  
C:Species: Influenza A virus  
C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 18-Sep-1998  
C:Accession: A29971  
R:Kida, H.; Shortridge, K.F.; Webster, R.G.  
Virology 162, 160-166, 1988  
A:Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China  
A:Reference number: A94370; M01D:88101364  
A:Accession: A29971  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <RID>  
A:Cross-references: GB:M19056; NID:G324208  
A>Note: the sequence in Genbank entry FLAHAB, release 106, (PID:G324209) differs from C:Genetics:  
A:Map position: segment 4  
C:Superfamily: Influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM>  
F:8,22,38,165,285,483/Binding site: carbohydrate (asn) (covalent) #status predicted  
F:114,466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7% Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1.4e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGFCAIAGFIENGEGMIDGMYG 24  
|||||

DB 330 GGFCAIAGFIENGEGMIDGMYG 352

#### RESULT 2

HMTIVS3  
hemagglutinin precursor - influenza A virus (strain A/swine/81/78) (fragment)

C:Species: Influenza A virus  
C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 18-Sep-1998

C:Accession: B29971  
R:Kida, H.; Shortridge, K.F.; Webster, R.G.

Virology 162, 160-166, 1988  
A:Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China

A:Reference number: A94370; M01D:88101364  
A:Accession: B29971

A:Molecule type: genomic RNA  
A:Residues: 1-550 <RID>

A:Cross-references: GB:M19057; NID:G324210  
A>Note: the sequence in Genbank entry FLAHAB, release 106, (PID:G324211) differs from C:Genetics:

A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:300-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14,466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%: Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%: Pred. No. 1,4e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFLENGEGMIDGMYG 24  
Db 330 GLFGAIAGFLENGEGMIDGMYG 352

RESULT 3  
HMTV77  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/5/77) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999  
C:Accession: A27813  
R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: A27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16737; NID:g324081; PIDN:AAA43145.1; PID:g324082  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14,466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%: Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%: Pred. No. 1,4e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFLENGEGMIDGMYG 24  
Db 330 GLFGAIAGFLENGEGMIDGMYG 352

RESULT 4  
HMTV80  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/8/80) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 18-Sep-1998  
C:Accession: B27813  
R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: B27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16738; NID:g324083  
A:Note: The translation in Fig. 2 is inconsistent with the nucleotide sequence in Fig. 1  
C:Genetics:

A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14,466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%: Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%: Pred. No. 1,4e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFLENGEGMIDGMYG 24  
Db 330 GLFGAIAGFLENGEGMIDGMYG 352

RESULT 5  
HMTV33  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/33/80) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999  
C:Accession: C27813  
R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: C27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16739; NID:g324085; PIDN:AAA43145.1; PID:g324086  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14,466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%: Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%: Pred. No. 1,4e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFLENGEGMIDGMYG 24  
Db 330 GLFGAIAGFLENGEGMIDGMYG 352

RESULT 6  
HMTV89  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/7/82) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999  
C:Accession: D27813  
R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: D27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16740; NID:g324087; PIDN:AAA43146.1; PID:g324088  
C:Genetics:  
A:Map position: segment 4

C:Superfamily: Influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;

Best Local Similarity 100.0%; Pred. No. 1.4e-10; Mismatches 0; Indels 0; Gaps 0;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAFGIENGWEGMIDGWTG 24

DB 330 GLFGAIAFGIENGWEGMIDGWTG 352

#### RESULT 7

hemagglutinin precursor - Influenza A virus (strain A/duck/Hokkaido/21/82) (fragment)

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: Influenza A virus

C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 18-Sep-1998

C:Accession: E27813

R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.

Virology 159, 109-119, 1987

A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.

A:Reference number: A94363; MUID:87265458

A:Accession: E27813

A:Molecule type: genomic RNA

A:Residues: 1-550 <KID>

A:Cross-references: GB:M16741; NID:9324089

C:Genetics:

A:Map position: segment 4

C:Superfamily: Influenza virus hemagglutinin

C:Keywords: glycoprotein; hemagglutinin; homotrimer

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>

F:520-536/Domain: transmembrane #status predicted <TM1>

F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

QY 2 GLFGAIAFGIENGWEGMIDGWTG 24

DB 330 GLFGAIAFGIENGWEGMIDGWTG 352

#### RESULT 9

hemagglutinin precursor - Influenza A virus (strain A/duck/Hokkaido/10/85) (fragment)

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: Influenza A virus

C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999

C:Accession: G27813

R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.

Virology 159, 109-119, 1987

A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.

A:Reference number: A94363; MUID:87265458

A:Accession: G27813

A:Molecule type: genomic RNA

A:Residues: 1-550 <KID>

A:Cross-references: GB:M16743; NID:9324093; PIDN:AAA43149.1; PID:9324094

C:Genetics:

A:Map position: segment 4

C:Superfamily: Influenza virus hemagglutinin

C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>

F:520-536/Domain: transmembrane #status predicted <TM1>

F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted



F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>  
F:8,22,38,165,285,483/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 2; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1,4e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGWTG 24  
DB 330 GLFGAIGFIENGWEGMIDGWTG 352

RESULT 11

hemagglutinin precursor - influenza A virus (strain A/goose/hong kong/10/76) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 20-Jun-2000  
C:Accession: J01154

R:Yaesuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.  
J. Gen. Virol. 72, 2007-2010, 1991

A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3

A:Reference number: J01153; MUID:91341491  
A:Accession: J01154

A:Molecule type: genomic RNA

A:Residues: 1-550 <YAS>

A:Cross-references: GB:D00930; NID:g221273; PIDN:BA00770.1; PID:g221274

A:Note: The authors translated the codon GGG for residue 218 as Glu

A:Note: residues 528-532 are not shown in this publication  
C:Superfamily: influenza virus hemagglutinin

C:Keywords: glycoprotein; homotrimer

F:1-338/Product: hemagglutinin HA1 #status predicted <HA1>

F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>

F:8,22,38,165,285,483/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 2; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1,4e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGWTG 24  
DB 330 GLFGAIGFIENGWEGMIDGWTG 352

RESULT 12

hemagglutinin precursor - influenza A virus (strain A/duck/hong kong/64/76) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 20-Jun-2000  
C:Accession: J01155

R:Yaesuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.  
J. Gen. Virol. 72, 2007-2010, 1991

A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3

A:Reference number: J01153; MUID:91341491  
A:Accession: J01155

A:Molecule type: genomic RNA

A:Residues: 1-550 <YAS>

A:Cross-references: GB:D00931; NID:g221277; PIDN:BA00771.1; PID:g221278

A:Note: the authors translated the codon GCG for residue 218 as Glu, GCC for residue 538

C:Superfamily: influenza virus hemagglutinin

C:Keywords: glycoprotein; homotrimer

F:1-338/Product: hemagglutinin HA1 #status predicted <HA1>

F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>

F:8,22,38,165,285,483/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 2; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1,4e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGWTG 24  
DB 330 GLFGAIGFIENGWEGMIDGWTG 352

RESULT 13

hemagglutinin precursor - influenza A virus

C:Species: influenza A virus

C:Date: 28-Feb-1981 #sequence\_revision 28-Feb-1981 #text\_change 22-Oct-1999

C:Accession: A93705; A93233; A04051; A93231; A94441

R:Both G.W.; Sleight, M.J.

Nucleic Acids Res. 8, 2561-2575, 1980

A:Title: Complete nucleotide sequence of the haemagglutinin gene from a human influ

A:Reference number: A93705; MUID:81053698

A:Accession: A93705

A:Molecule type: genomic RNA

A:Residues: 1-566 <BOT>

A:Cross-references: GB:V01103

A:Experimental source: strain A/NT/60/68/29C

A:Note: human influenza strain A/NT/60/68/29C is a laboratory-isolated variant of A,

R:Dopheide, T.A.; Ward, C.W.

FEBS Lett. 110, 181-183, 1980

A:Title: The disulphide bonds of a Hong Kong influenza virus hemagglutinin.

A:Reference number: A91276; MUID:80179105

A:Contents: annotation; disulfide bonds

R:Gethling, M.J.; Bye, J.; Skene, J.; Waterfield, M.

Nature 287, 301-306, 1980

A:Title: Cloning and DNA sequence of double-stranded copies of haemagglutinin genes

A:Reference number: A93233; MUID:81030852

A:Accession: A93233

A:Molecule type: genomic RNA

A:Residues: 1-24, 'S', 26, 'D', 28-159, 'G', 161-197, 'I', 199-241, 'L', 243-249 <GET>

A:Experimental source: strain X-31[H3]

C:Superfamily: influenza virus hemagglutinin

C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>

F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>

F:566-552/Domain: transmembrane #status predicted <TM1>

F:30-482,68-293,80-92,155-489,297-321/Disulfide bonds: #status experimental

F:555,562,565/Binding site: palmitate (cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 1,4e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGWTG 24  
DB 346 GLFGAIGFIENGWEGMIDGWTG 368

RESULT 14

hemagglutinin precursor - influenza A virus (strain A/Atch/2/68)

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: influenza A virus

C:Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 16-Jul-1999

C:Accession: A93231; A04051

R:Verhoeven, M.; Pang, R.; Min Jou, W.; Devos, R.; Huybrecock, D.; Saman, E.; Pieri

Nature 286, 771-776, 1980

A:Title: Antigenic drift between the haemagglutinin of the Hong Kong influenza stra

A:Reference number: A93231; MUID:80254693

A:Accession: A93231

A:Molecule type: genomic RNA

A:Residues: 1-566 <VER>

A:Cross-references: GB:J02090; NID:g324131; PIDN:AAA3178.1; PID:g324132

C:Superfamily: influenza virus hemagglutinin

C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>

F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>  
F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match

93.7%; Score 134; DB 1; Length 566;

Best Local Similarity 100.0%; Pred. No. 1,4e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGMYG 24  
|||||

Db 346 GLFGAIGFIENGWEGMIDGMYG 368

RESULT 15

HM1VHM

hemagglutinin precursor - influenza A virus (strain A/Mem/102/72)

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: Influenza A virus

C>Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 31-Mar-2000

C:Accession: A94441; A04051

R:Stoib, M.J.; Both, G.W.; Browlee, G.G.; Bender, V.J.; Moss, B.A.

in Structure and Variation in Influenza Virus, Laver, G., and Air, G., eds., pp.69-79, F

A:Title: The haemagglutinin gene of influenza A virus: nucleotide sequence analysis of c

A:Reference number: A94441

A:Accession: A94441

A:Molecule type: genomic RNA

A:Residues: 1-566 <SIDE>

C:Superfamily: Influenza virus hemagglutinin

C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>

F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>

F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match

93.7%; Score 134; DB 1; Length 566;

Best Local Similarity 100.0%; Pred. No. 1,4e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGMYG 24  
|||||

Db 346 GLFGAIGFIENGWEGMIDGMYG 368

Search completed: July 1, 2002, 06:28:11  
Job time: 713 sec

33	131	91.6	565	1	HEMA_IAMI	P15658	influenza a
----	-----	------	-----	---	-----------	--------	-------------

Best Local Similarity 100.0%; Pred. No. 1.1e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGEGMIDGWTG 24  
|||||  
DB 330 GLFGAIGFIENGEGMIDGWTG 352

RESULT 2  
HEMA\_IADH2 STANDARD: PRT: 550 AA.  
ID HEMA\_IADH2  
AC P12583; Q84011;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
Hemagglutinin HA2 chain] (Fragment).  
GN HA.  
OS Influenza A virus (strain A/Duck/Hokkaido/8/80).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
CC Influenza virus A and B group; Influenza A viruses;  
CC Influenza A virus.  
OX NCBI\_TaxID=11359;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87265458; PubMed=2440178;  
RA Kida H., Kawoka Y., Naeve C.W., Webster R.G.;  
RT "Antigenic and genetic conservation of H3 influenza virus in wild  
ducks.";  
RT Virology 159:109-119(1987).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
CC EMBL: M16738; AAA43144.1; -  
DR PIR: B27813; HMIV80.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR ProDom: PD000225; Hemagglutn; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON\_TER 1 1  
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
FT CARBOHYD 330 550 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 137 137 K -> N (IN PIR DATA BANK).  
SQ SEQUENCE 550 AA; 61659 MW; A107023AC9CC353 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1,1e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGEGMIDGWTG 24  
|||||  
DB 330 GLFGAIGFIENGEGMIDGWTG 352

RESULT 3  
HEMA\_IADH3

ID HEMA\_IADH3 STANDARD: PRT: 550 AA.  
AC P12584; Q84012; Q89793;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
Hemagglutinin HA2 chain] (Fragment).  
GN HA.  
OS Influenza A virus (strain A/Duck/Hokkaido/33/80).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
CC Influenza virus A and B group; Influenza A viruses;  
CC Influenza A virus.  
OX NCBI\_TaxID=11359;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87265458; PubMed=2440178;  
RA Kida H., Kawoka Y., Naeve C.W., Webster R.G.;  
RT "Antigenic and genetic conservation of H3 influenza virus in wild  
ducks.";  
RT Virology 159:109-119(1987).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
CC EMBL: M16739; AAA43145.1; -  
DR PIR: C27813; HMIY33.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR ProDom: PD000225; Hemagglutn; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON\_TER 1 1  
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
FT CARBOHYD 330 550 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 550 AA; 61577 MW; 6C30BF67CFDCB7DE CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1,1e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGEGMIDGWTG 24  
|||||  
DB 330 GLFGAIGFIENGEGMIDGWTG 352

RESULT 4  
HEMA\_IADH4 STANDARD: PRT: 550 AA.  
ID HEMA\_IADH4  
AC P12585; Q84013; Q84014;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
Hemagglutinin HA2 chain] (Fragment).  
GN HA.  
OS Influenza A virus (strain A/Duck/Hokkaido/77/82).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

```
CC Influenza virus A and B group; Influenza A viruses;
CC Influenza A virus.
CC NCBI_TaxID=11360;
RN
  [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaoaka Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
  ducks.";
RL virology 159:109-119(1987).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
  CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
  (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch.)
CC -----
CC
CC EMBL: M16740; AAA43146.1; -.
CC PIR: D27813; HMTV89.
CC InterPro: IPR001364; Hemagglutn.
CC Pfam: PF00509; Hemagglutinin; 1.
CC ProDom: PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein.
CC NON_TER 1 1
CC CHAIN 1 328
CC CHAIN 330 550
CC CARBOHYD 8 8
CC CARBOHYD 22 22
CC CARBOHYD 38 38
CC CARBOHYD 165 165
CC CARBOHYD 285 285
CC CARBOHYD 483 483
CC SEQUENCE 550 AA; 61664 MW; A16B2CF8CBBBD9D0 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1,1e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDWYG 24
Db 330 GLFGAIGFIENGWEGMIDWYG 352
|||||
RESULT 5
HEMA_IADH5 STANDARD; PRT; 550 AA.
ID HEMA_IADH5
AC P12586; Q84015; Q84016;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;
  Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/21/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11361;
RN
  [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaoaka Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
  ducks.";
RL virology 159:109-119(1987).
```

```
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
  CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
  (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch.)
CC -----
CC
CC EMBL: M16741; AAA43147.1; -.
CC PIR: E27813; HMTV21.
CC HSP: F03437; 5HMG.
CC InterPro: IPR001364; Hemagglutn.
CC Pfam: PF00509; Hemagglutinin; 1.
CC ProDom: PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein.
CC NON_TER 1 1
CC CHAIN 1 328
CC CHAIN 330 550
CC CARBOHYD 7 7
CC CARBOHYD 8 8
CC CARBOHYD 22 22
CC CARBOHYD 38 38
CC CARBOHYD 165 165
CC CARBOHYD 285 285
CC CARBOHYD 483 483
CC CARBOHYD 178 178
CC CONFLICT 368 388
CC SEQUENCE 550 AA; 61856 MW; 48401C67A15BF8C CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1,1e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDWYG 24
Db 330 GLFGAIGFIENGWEGMIDWYG 352
|||||
RESULT 6
HEMA_IADH6 STANDARD; PRT; 550 AA.
ID HEMA_IADH6
AC P12587; Q84017;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;
  Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/9/85).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11362;
RN
  [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaoaka Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
  ducks.";
RL virology 159:109-119(1987).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
  CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
  (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M16742; AAA43148.1; -  
CC PIR: F27813; HMTV98.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR ProDom: PD000225; Hemagglutn; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON\_TER 1 328  
FT CHAIN 1 328  
FT CARBOHYD 330 550  
FT CARBOHYD 22 22  
FT CARBOHYD 38 38  
FT CARBOHYD 165 165  
FT CARBOHYD 285 285  
FT CARBOHYD 483 483  
FT CONFLICT 8 8  
SQ SEQUENCE 550 AA; 61711 MW; 67BCD85F44736CFC CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1,1e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLEGAAGFENGEGMIDWYG 24  
DQ 330 GLEGAAGFENGEGMIDWYG 352

RESULT 7

HEMA\_IADH7 STANDARD; PRT; 550 AA.

AC P12588; Q84018; Q89470;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;

DE Hemagglutinin HA2 chain] (Fragment).

GN HA.

OS Influenza A virus (strain A/Duck/Hokkaido/10/85).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza virus A and B group; Influenza A viruses;

CC Influenza A virus.

OX NCBI\_TaxID=11363;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=87265458; PubMed=2440178;

RA Kida H., Kawaka Y., Naeve C.W., Webster R.G.;

RT "Antigenic and genetic conservation of H3 influenza virus in wild ducks";

RL Virology 159:109-119(1987).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CC -1- CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS

CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR ProDom: PD000225; Hemagglutn; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON\_TER 1 328  
FT CHAIN 1 328  
FT CARBOHYD 330 550  
FT CARBOHYD 22 22  
FT CARBOHYD 38 38  
FT CARBOHYD 165 165  
FT CARBOHYD 285 285  
FT CARBOHYD 483 483  
FT CONFLICT 8 8  
SQ SEQUENCE 550 AA; 61761 MW; 6ER81793281D53EB CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1,1e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLEGAAGFENGEGMIDWYG 24  
DQ 330 GLEGAAGFENGEGMIDWYG 352

RESULT 8

HEMA\_IADH7 STANDARD; PRT; 550 AA.

AC P43257;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;

DE Hemagglutinin HA2 chain] (Fragment).

GN HA.

OS Influenza A virus (strain A/Duck/Hong Kong/7/75).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza virus A and B group; Influenza A viruses;

CC Influenza A virus.

OX NCBI\_TaxID=11364;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=91341491; PubMed=1875195;

RA Yasuda U., Shortridge K.F., Shimizu Y., Kida H.;

RT "Molecular evidence for a role of domestic ducks in the introduction

RT of avian H3 influenza viruses to pigs in southern China, where the

RT A/Hong Kong/68 (H3N2) strain emerged";

RL J. Gen. Virol. 72:2007-2010(1991).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CC -1- CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS

CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: D00929; BAA00769.1; -

CC HSSP: P03437; SHMG.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR ProDom: PD000225; Hemagglutn; 1.

KW Envelope protein; Hemagglutinin; Glycoprotein.

FT NON\_TER 1 328

FT CHAIN 1 328

FT CARBOHYD 330 550

FT CARBOHYD 22 22

FT CARBOHYD 38 38

FT CARBOHYD 165 165

FT CARBOHYD 285 285

FT CARBOHYD 483 483

FT CONFLICT 8 8

FT CARBOHYD 22 22

FT CARBOHYD 8 8

FT CARBOHYD 22 22

FT CARBOHYD 8 8

FT CARBOHYD 22 22

FT CARBOHYD 8 8

FT CARBOHYD 22 22

FT CARBOHYD 8 8

FT CARBOHYD 22 22

FT CARBOHYD 8 8

FT CARBOHYD 22 22

FT CARBOHYD 8 8

FT CARBOHYD 22 22

FT CARBOHYD 8 8

FT CARBOHYD 22 22

FT CARBOHYD 8 8

FT CARBOHYD 22 22

FT CARBOHYD 8 8

FT CARBOHYD 22 22

FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 550 AA; 61549 MW; 864639829FE1B9A CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1.1e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLEGAIGFIENGWEGMIDWYG 24  
Db 330 GLEGAIGFIENGWEGMIDWYG 352

RESULT 9  
HEMA\_IADHL STANDARD; PRT; 550 AA.  
ID HEMA\_IADHL  
AC P43258;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;  
DE Hemagglutinin HA2 chain] (Fragment).  
GN HA.  
OS Influenza A virus (strain A/Duck/Hong Kong/64/76).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses;  
OC Influenza A virus.  
OX NCBI\_TaxID=45412;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91341491; PubMed=1875195;  
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;  
RT "Molecular evidence for a role of domestic ducks in the introduction  
RT of avian H3 influenza viruses to pigs in southern China, where the  
RT A/hong kong/58 (H3N2) strain emerged."  
RL J. Gen. Virol. 72:2007-2010(1991).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: D00931; BA00771.1; -  
CC HSSP: P03437; 5HMG.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR ProDom: PD000225; Hemagglutn; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON\_TER 1 1  
FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.  
FT CHAIN 1 330 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 550 AA; 61718 MW; A35156789E4B9A CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1.1e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GLEGAIGFIENGWEGMIDWYG 24  
Db 330 GLEGAIGFIENGWEGMIDWYG 352

RESULT 10  
HEMA\_IAGHK STANDARD; PRT; 550 AA.  
ID HEMA\_IAGHK  
AC P43260;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;  
DE Hemagglutinin HA2 chain] (Fragment).  
GN HA.  
OS Influenza A virus (strain A/Goose/Hong Kong/10/76).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses;  
OC Influenza A virus.  
OX NCBI\_TaxID=45414;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91341491; PubMed=1875195;  
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;  
RT "Molecular evidence for a role of domestic ducks in the introduction  
RT of avian H3 influenza viruses to pigs in southern China, where the  
RT A/hong kong/58 (H3N2) strain emerged."  
RL J. Gen. Virol. 72:2007-2010(1991).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: D00930; BA00770.1; -  
CC HSSP: P03437; 5HMG.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR ProDom: PD000225; Hemagglutn; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON\_TER 1 1  
FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.  
FT CHAIN 1 330 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 550 AA; 61676 MW; 9A1D094DA28BBD2 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1.1e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLEGAIGFIENGWEGMIDWYG 24  
Db 330 GLEGAIGFIENGWEGMIDWYG 352

RESULT 11  
HEMA\_IAGH2

```

ID HEMA_IAZH2 STANDARD: PRT: 550 AA.
AC P1133: 084019; 084020.
DT 01-JUL-1988 (Rel. 11, Created)
DT 13-JUL-1988 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin [contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Swine/Hong Kong/81/78).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC NCBI_TaxID=11497;
OX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88101364; PubMed=3336940;
RA Kida H., Shortridge K.F., Webster R.G.;
RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China."
RL Virology 162:160-166(1988).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19057; AAA43212.1; .
DR PIR: B29971; HMY53.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutn; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Hemagglutinin; Envelope protein; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 550 AA; 61437 MW; 1F2A7E758C531CE8 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.1e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGEGMIDGWG 24
DB 330 GLFGAIGFIENGEGMIDGWG 352

RESULT 12
HEMA_IAZH3 STANDARD: PRT: 550 AA.
AC P1134: 084025; 084026;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin [contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Swine/Hong Kong/126/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

```

```

OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OC NCBI_TaxID=11498;
OX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88101364; PubMed=3336940;
RA Kida H., Shortridge K.F., Webster R.G.;
RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China."
RL Virology 162:160-166(1988).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19056; AAA43211.1; AUL_TERM.
DR PIR: A29971; HMY52.
DR HSSP: P03437; 2HWG.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutn; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Hemagglutinin; Envelope protein; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 550 AA; 61580 MW; 991F6DB8C02F24F2 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.1e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGEGMIDGWG 24
DB 330 GLFGAIGFIENGEGMIDGWG 352

RESULT 13
HEMA_IAAIC STANDARD: PRT: 566 AA.
AC P03437;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Alchi/2/68).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=150147;
RP SEQUENCE FROM N.A.
RX MEDLINE=80254693; PubMed=7402351;
RA Verhoeven M., Fang R., Min Jou W., Devos R., Huylenbroeck D., Saman E., Fiers W.;
RT "Antigenic drift between the haemagglutinin of the Hong Kong

```





Query Match 93.7%; Score 134; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 1.1e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

FT STRAND 310 311  
FT STRAND 318 320  
FT STRAND 323 324  
FT STRAND 331 333  
FT STRAND 337 337  
FT STRAND 347 348  
FT TURN 350 350  
FT STRAND 351 351  
FT TURN 352 354  
FT STRAND 355 355  
FT STRAND 359 359  
FT TURN 360 361  
FT STRAND 367 372  
FT TURN 374 375  
FT STRAND 378 382  
FT HELIX 383 401  
FT STRAND 406 407  
FT HELIX 421 474  
FT STRAND 475 477  
FT STRAND 482 485  
FT HELIX 491 498  
FT TURN 499 500  
FT HELIX 505 515  
SQ SEQUENCE 566 AA; 63415 MW; E395659C23CAFECA CRC64;

RESULT 14  
HEMA\_IAD3 STANDARD; PRT; 566 AA.  
ID HEMA\_IAD3  
AC P26134;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;  
DE Hemagglutinin HAZ chain].  
GN HA.  
OS Influenza A virus (strain A/Duck/Alberta/79/76).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses;  
OC Influenza A virus.  
OX NCBI\_TaxID=11348;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92114135; PubMed=1731092;  
RA Bean W.J., Schnell M., Katz J., Kawoka Y., Naeye C., Gorman O.,  
RA Webster R.G.;  
RT "Evolution of the H3 influenza virus hemagglutinin from human and  
RT nonhuman hosts".  
RU J. Virol. 66:1129-1138(1992).  
CC - FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC - SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC - SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M73771; -; NOT\_ANNOTATED\_CDS.  
DR

DR HSP; P03437; SHMC.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin.1.  
DR PRINTS: PR00329; HEMAGGLUTN1.2.  
DR ProDom: PD000225; Hemagglutn.1.  
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.  
FT SIGNAL 1 16  
FT CHAIN 17 344 HEMAGGLUTININ HAI CHAIN.  
FT CHAIN 346 566 HEMAGGLUTININ HAZ CHAIN.  
FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 566 AA; 63534 MW; FE19AB6F9415B89 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 1.1e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 GLFGAIGFIENGWEGMIDGMYG 24  
D 346 GLFGAIGFIENGWEGMIDGMYG 368

RESULT 15  
HEMA\_IAD3 STANDARD; PRT; 566 AA.  
ID HEMA\_IAD3  
AC P03442;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;  
DE Hemagglutinin HAZ chain].  
GN HA.  
OS Influenza A virus (strain A/Duck/Ukraine/1/63).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses;  
OC Influenza A virus.  
OX NCBI\_TaxID=11374;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=82025542; PubMed=6169439;  
RA Fang R., Min Jou W., Huybrecock D., Devos R., Fiers W.;  
RA "Complete structure of A/Duck/Ukraine/63 influenza hemagglutinin  
RT gene: animal virus as progenitor of human H3 Hong Kong 1968 influenza  
RT hemagglutinin".  
RU Cell 25:315-323(1981).  
CC - FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC - SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC - SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: V01087; CAA24271.1; -.  
DR PIR: A04053; HMIYDU.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin.1.  
DR PRINTS: PR00329; HEMAGGLUTN1.2.  
DR ProDom: PD000225; Hemagglutn.1.  
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.  
FT SIGNAL 1 16

FT	CHAIN	17	344	HEMAGGLUTININ HA1 CHAIN.
FT	CHAIN	346	566	HEMAGGLUTININ HA2 CHAIN.
FT	CARBOHYD	24	24	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	38	38	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	54	54	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	97	97	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	181	181	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	301	301	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	499	499	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	566 AA;	63530 MM;	E70F87F0AE1178F4 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 566;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-10;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIENGWEGMIDGWS 24  
 DB 346 GLFGAIAGFIENGWEGMIDGWS 368

Search completed: July 1, 2002, 06:31:35  
 Job time: 672 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:30:57 ; Search time 83.98 Seconds  
(without alignments)  
49.439 Million cell updates/sec

Title: US-09-461-684-3  
Perfect score: 143  
Sequence: 1 CGLFGAIGFIENGMEGMDWYG 24

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 segs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP:REMBL\_19:\*  
2: SP:Archaea:\*  
3: SP:Bacteria:\*  
4: SP:Fungi:\*  
5: SP:Human:\*  
6: SP:Invertebrate:\*  
7: SP:Mammal:\*  
8: SP:pmhc:\*  
9: SP:organelle:\*  
10: SP:Phage:\*  
11: SP:Plant:\*  
12: SP:rodent:\*  
13: SP:virus:\*  
14: SP:vertebrate:\*  
15: SP:unclassified:\*  
16: SP:Virus:\*  
17: SP:Bacteriap:\*  
17: SP:archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	134	93.7	566	12	Q9DHG0 Influenza a
2	134	93.7	566	12	Q67125 Influenza a
3	134	93.7	566	12	Q67126 Influenza a
4	134	93.7	566	12	Q67132 Influenza a
5	134	93.7	566	12	Q98052 Influenza a
6	134	93.7	566	12	Q91MA7 Influenza a
7	134	93.7	566	12	Q91MA7 Influenza a
8	133	93.0	301	12	Q9DXE3 Influenza a
9	133	92.3	550	12	Q9DXE3 Influenza a
10	132	92.3	550	12	Q82753 Influenza a
11	132	92.3	550	12	Q82498 Influenza a
12	132	92.3	566	12	Q82496 Influenza a
13	132	92.3	571	12	Q03909 Influenza a
14	131	91.6	109	12	Q67050 Influenza a
15	131	91.6	109	12	Q67051 Influenza a
16	131	91.6	109	12	Q67052 Influenza a

RESULT	ID	Q9DHG0	PRELIMINARY:	PRT:	566 AA.	ALIGNMENTS
17	131	91.6	109	12	Q67053	Influenza a
18	131	91.6	362	12	Q82513	Influenza a
19	131	91.6	362	12	Q82517	Influenza a
20	131	91.6	362	12	Q84174	Influenza a
21	131	91.6	362	12	Q9QKD3	Influenza a
22	131	91.6	362	12	Q9QKD2	Influenza a
23	131	91.6	362	12	Q9QKD1	Influenza a
24	131	91.6	365	12	Q9DL25	Influenza a
25	131	91.6	367	12	Q9DL22	Influenza a
26	131	91.6	368	12	Q9DL29	Influenza a
27	131	91.6	369	12	Q9DL26	Influenza a
28	131	91.6	369	12	Q9DL06	Influenza a
29	131	91.6	369	12	P87689	Influenza a
30	131	91.6	371	12	Q9DL24	Influenza a
31	131	91.6	371	12	P87685	Influenza a
32	131	91.6	373	12	Q9DL20	Influenza a
33	131	91.6	374	12	Q9DL21	Influenza a
34	131	91.6	375	12	Q9DL27	Influenza a
35	131	91.6	375	12	Q9DL05	Influenza a
36	131	91.6	376	12	Q9DL30	Influenza a
37	131	91.6	376	12	Q9DL04	Influenza a
38	131	91.6	377	12	Q9E7P5	Influenza a
39	131	91.6	382	12	Q9DL03	Influenza a
40	131	91.6	408	12	Q9E7P5	Influenza a
41	131	91.6	409	12	Q9Q0L5	Influenza a
42	131	91.6	416	12	Q9Q0C4	Influenza a
43	131	91.6	429	12	Q9Q0L4	Influenza a
44	131	91.6	438	12	Q9Q0L3	Influenza a
45	131	91.6	467	12	Q9E312	Influenza a

#### ALIGNMENTS

RESULT 1  
Q9DHG0  
AC Q9DHG0: PRELIMINARY: PRT: 566 AA.  
DT 01-MAR-2001 (TREMELrel. 16, Created)  
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)  
DE HAEMAGGLUTININ PRECURSOR.  
OS Influenza A virus H3N2.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
OC NCBI\_taxid=41857;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CLONE 7A (H3N2);  
RA Moshin M.A., Morris S.J., Smith H., Sweet C.;  
RT "Influenza virus-induced apoptosis: a dual role for viral  
neuraminidase.";  
RL Submitted (Jun-2000) to the EMBL/GenBank/DBJ databases.  
CC - FUNCTION: HAEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC - SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC - SIMILARITY: BELONGS TO THE INFLUENZA HAEMAGGLUTININ FAMILY.  
DR EMBL: AJ289703; CAC18525.1;  
DR HSRP: P03437; 2YIU.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR PRINTS: PR00329; HAEMAGGLUTN12.  
DR ProDom: PD000225; Hemagglutn; 1.  
KW Envelope protein; Glycoprotein; Hemagglutinin; Signal.  
FT SIGNAL 1  
SQ SEQUENCE 566 AA: 63356 MW: 0BA68192300F72F CRC64;

Query Match 93.7%; Score 134; DB 12; Length 566;  
Best Local Similarity 100.0%; Pred. No. 2.2e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGEGMIDGWTG 24  
|||||  
DB 346 GLFGAIGFIENGEGMIDGWTG 368

RESULT 2  
067125 PRELIMINARY; PRT; 566 AA.  
AC 067125;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE HEMAGGLUTININ.  
GN HA.  
OS Influenza A virus.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCBI\_TaxID=11320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/SEAL/MA/3911/92;  
RX MEDLINE=95146951; PubMed=7844533;  
RA Callan R.J., Early G., Kida H., Hinshaw V.S.;  
RT "The appearance of H3 influenza viruses in seals.";  
RL J. Gen. Virol. 76:199-203(1995).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL; L31949; AAA64229.1; -.  
DR HSSP; P03437; ZVIU.  
DR InterPro; IPR001364; Hemagglutn.  
DR Pfam; PF00509; Hemagglutinin; 1.  
DR PRINTS; PR00329; HEMAGGLUTIN12.  
DR PRODOM; PD000225; Hemagglutn; 1.  
KW Envelope protein; Glycoprotein; Hemagglutinin.  
SQ SEQUENCE 566 AA; 63456 MW; AE556302A9EB99F CRC64;

Query Match 93.7%; Score 134; DB 12; Length 566;  
Best Local Similarity 100.0%; Pred. No. 2.2e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGEGMIDGWTG 24  
|||||  
DB 346 GLFGAIGFIENGEGMIDGWTG 368

RESULT 3  
067126 PRELIMINARY; PRT; 566 AA.  
AC 067126;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE HEMAGGLUTININ.  
GN HA.  
OS Influenza A virus.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCBI\_TaxID=11320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/SEAL/MA/3984/92;  
RX MEDLINE=95146951; PubMed=7844533;  
RA Callan R.J., Early G., Kida H., Hinshaw V.S.;  
RT "The appearance of H3 influenza viruses in seals.";  
RL J. Gen. Virol. 76:199-203(1995).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL; M55059; AAA43239.1; -.  
DR HSSP; P03437; IHCE.  
DR InterPro; IPR001364; Hemagglutn.  
DR Pfam; PF00509; Hemagglutinin; 1.  
DR PRINTS; PR00329; HEMAGGLUTIN12.  
DR PRODOM; PD000225; Hemagglutn; 1.  
KW Envelope protein; Glycoprotein; Hemagglutinin.  
FT CHAIN 1 344 HEMAGGLUTININ.  
FT CHAIN 346 566 NEURAMINIDASE.  
SQ SEQUENCE 566 AA; 63441 MW; ESD1B97DF96FECA CRC64;

Query Match 93.7%; Score 134; DB 12; Length 566;  
Best Local Similarity 100.0%; Pred. No. 2.2e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGEGMIDGWTG 24  
|||||  
DB 346 GLFGAIGFIENGEGMIDGWTG 368

RESULT 5  
098052 PRELIMINARY; PRT; 566 AA.  
AC 098052;  
DT 01-FEB-1997 (TREMBlrel. 02, Created)

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL; L32024; AAA64228.1; -.  
DR HSSP; P03437; ZVIU.  
DR InterPro; IPR001364; Hemagglutn.  
DR Pfam; PF00509; Hemagglutinin; 1.  
DR PRINTS; PR00329; HEMAGGLUTIN12.  
DR PRODOM; PD000225; Hemagglutn; 1.  
KW Envelope protein; Glycoprotein; Hemagglutinin.  
SQ SEQUENCE 566 AA; 63441 MW; 590576CB4CEE7D08 CRC64;

Query Match 93.7%; Score 134; DB 12; Length 566;  
Best Local Similarity 100.0%; Pred. No. 2.2e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGEGMIDGWTG 24  
|||||  
DB 346 GLFGAIGFIENGEGMIDGWTG 368

RESULT 4  
067132 PRELIMINARY; PRT; 566 AA.  
AC 067132;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE HEMAGGLUTININ.  
GN HA.  
OS Influenza A virus (strain A/Alchi/2/68).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCBI\_TaxID=150147;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/ALCHI/2/68;  
RA Min J.W., Verhoeven M., Fang R.-X., Devos R., Huybreoek D.,  
RA Fiers W.;  
RT "Shift and drift in influenza viruses."; (in) Carille M.J., Collins J.F., Moseley B.E. B. (eds.);  
RL SYMPOSIUM OF THE SOCIETY FOR GENERAL MICROBIOLOGY, pp.285-311,  
CC Cambridge University Press, New York (1991).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL; M55059; AAA43239.1; -.  
DR HSSP; P03437; IHCE.  
DR InterPro; IPR001364; Hemagglutn.  
DR Pfam; PF00509; Hemagglutinin; 1.  
DR PRINTS; PR00329; HEMAGGLUTIN12.  
DR PRODOM; PD000225; Hemagglutn; 1.  
KW Envelope protein; Glycoprotein; Hemagglutinin.  
FT CHAIN 1 344 HEMAGGLUTININ.  
FT CHAIN 346 566 NEURAMINIDASE.  
SQ SEQUENCE 566 AA; 63441 MW; ESD1B97DF96FECA CRC64;

Query Match 93.7%; Score 134; DB 12; Length 566;  
Best Local Similarity 100.0%; Pred. No. 2.2e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGEGMIDGWTG 24  
|||||  
DB 346 GLFGAIGFIENGEGMIDGWTG 368

RESULT 5  
098052 PRELIMINARY; PRT; 566 AA.  
AC 098052;  
DT 01-FEB-1997 (TREMBlrel. 02, Created)

```

DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HEMAGGLUTININ PRECURSOR (FRAGMENT).
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=61053698; PubMed=6253883;
RA Both G.W., Sleight M.J.;
RT "Complete nucleotide sequence of the haemagglutinin gene from a human
RT Influenza virus of the Hong Kong subtype."
RL Nucleic Acids Res. 8:2561-2575(1980).
RN [2]
RP SEQUENCE OF 17-344 FROM N.A.
RX MEDLINE=61194918; PubMed=6164798;
RA Sleight M.J., Both G.W., Underwood P.A., Bender V.J.;
RT "Antigenic drift in the hemagglutinin of the Hong Kong influenza
RT subtype: Correlation of amino acid changes with alterations in viral
RT antigenicity."
RL J. Virol. 37:845-853(1981).
RN [3]
RP SEQUENCE OF 17-566 FROM N.A.
RX MEDLINE=62033276; PubMed=6169843;
RA Both G.W., Sleight M.J.;
RT "Conservation and variation in the hemagglutinins of Hong Kong subtype
RT Influenza viruses during antigenic drift."
RL J. Virol. 39:845-853(1981).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL: J02135; AAA3189.1; -
DR HSSP: P03437; 1HGE.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin.1.
DR PRINTS: PR00329; HEMAGGLUTIN12.
DR PRODOM: PD000225; Hemagglutn.1.
KW Envelope protein; Glycoprotein; Hemagglutinin; Signal.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 344 POTENTIAL.
FT CHAIN 346 566 POTENTIAL.
SQ SEQUENCE 566 AA; 63414 MW; C447FD465BE4FCF9 CRC64;

```

Query Match 93.7%; Score 134; DB 12; Length 566;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-10;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 2 GLFGAIGFIENGWEGMIDGWY 24
DB 346 GLFGAIGFIENGWEGMIDGWY 368

```

RESULT 6  
 Q91MA7 PRELIMINARY; PRT; 566 AA.

AC Q91MA7;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HEMAGGLUTININ (FRAGMENT).  
 OS Influenza A virus (A/Hong Kong/1/68(H3N2)).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
 OX NCBI\_TaxID=108859;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A/HONG KONG/1/68(H3N2);  
 RX MEDLINE=21287244; PubMed=11371620;  
 RT "Pattern of mutation in the genome of influenza A virus on adaptation
 RT to increased virulence in the mouse lung: Identification of functional
 RT themes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).  
 DR EMBL: AF348179; AAK5172.1; -  
 DR EMBL: AF348177; AAK51719.1; -  
 DR EMBL: AF348178; AAK51720.1; -  
 SQ SEQUENCE 566 AA; 63530 MW; 7CB9F5BAF1E6E9F4 CRC64;

```

RT "Pattern of mutation in the genome of influenza A virus on adaptation
RT to increased virulence in the mouse lung: Identification of functional
RT themes."
RT Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).
DR EMBL: AF348176; AAK51718.1; -
SQ SEQUENCE 566 AA; 63387 MW; 01B0D465BE15BE1 CRC64;

```

Query Match 93.7%; Score 134; DB 12; Length 566;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-10;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 2 GLFGAIGFIENGWEGMIDGWY 24
DB 346 GLFGAIGFIENGWEGMIDGWY 368

```

RESULT 7  
 Q91OM5 PRELIMINARY; PRT; 566 AA.

AC Q91OM5;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HEMAGGLUTININ.  
 OS Influenza A virus (A/Hong Kong/1/68(H3N2)).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
 OX NCBI\_TaxID=108859;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A/HONG KONG/1/68(H3N2);  
 RX MEDLINE=21287244; PubMed=11371620;  
 RT "Pattern of mutation in the genome of influenza A virus on adaptation
 RT to increased virulence in the mouse lung: Identification of functional
 RT themes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).  
 DR EMBL: AF348179; AAK5172.1; -  
 DR EMBL: AF348177; AAK51719.1; -  
 DR EMBL: AF348178; AAK51720.1; -  
 SQ SEQUENCE 566 AA; 63530 MW; 7CB9F5BAF1E6E9F4 CRC64;

```

QY 2 GLFGAIGFIENGWEGMIDGWY 24
DB 346 GLFGAIGFIENGWEGMIDGWY 368

```

RESULT 8  
 Q9DXE3 PRELIMINARY; PRT; 301 AA.

AC Q9DXE3;  
 DT 01-MAR-2001 (TREMBLrel. 15, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HEMAGGLUTININ (FRAGMENT).  
 OS Influenza A virus (A/Shorebird/Taiwan/31-4/99).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
 OX NCBI\_TaxID=140665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A/SHOREBIRD/TAIWAN/31-4/99;  
 RA Lee M.S., Cheng P.C., Shien J.H., Cheng M.C., Lee L.H., Shieh H.K.;  
 RT "Identification and subtyping of avian influenza virus by reverse
 RT transcription-polymerase chain reaction.";  
 RL Submitted (Oct-2000) to the EMBL/Genbank/DBJ databases.

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL: AF11750; AAC33016.1; -  
DR InterPro: IPR001364; Hemagglutn.  
DR PRINTS: PR00329; HEMAGGLUTN12.  
DR ProDom: PD000225; Hemagglutn.1.  
KW Envelope protein; Glycoprotein; Hemagglutinin.  
FT NON-TER 1  
FT CHAIN 1  
SQ SEQUENCE 301 AA; 32701 MW; 62A03758B764D57 CRC64;

Query Match 93.0%; Score 133; DB 12; Length 301;  
Best Local Similarity 95.7%; Pred. No. 1.5e-10;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFENGWEGMIDGWG 24  
DB 250 GLFGAIGFENGWEGMIDGWG 272

RESULT 9  
ID 082498 PRELIMINARY; PRT; 550 AA.  
AC 082498;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE HEMAGGLUTININS HA1 AND HA2 (FRAGMENT).  
OS Influenza A virus.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCBI\_TaxID=11320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/PHILIPPINES/2/82/BS (H3N2);  
RA Hartley C.A., Ward A.C., Anders E.M.;  
RT "Virulence of Influenza virus for mice is associated with loss of  
RT oligosaccharide from the hemagglutinin molecule."  
RL Submitted (Apr-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL: U08859; AAA16782.1; -  
DR HSSP; P03437; 2YIU.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin.1.  
DR PRINTS: PR00329; HEMAGGLUTN12.  
DR ProDom: PD000225; Hemagglutn.1.  
KW Envelope protein; Glycoprotein; Hemagglutinin.  
FT NON-TER 1  
FT CHAIN 1  
FT CHAIN 1  
FT CHAIN 1  
SQ SEQUENCE 550 AA; 61772 MW; 50B5D2B5BF11FD8 CRC64;

Query Match 92.3%; Score 132; DB 12; Length 550;  
Best Local Similarity 95.7%; Pred. No. 4e-10;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFENGWEGMIDGWG 24  
DB 330 GLFGAIGFENGWEGMIDGWG 352

RESULT 10  
O82753 PRELIMINARY; PRT; 550 AA.  
ID 082753

AC 082753;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE HAEMAGGLUTININ (FRAGMENT).  
OS Influenza virus.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC unclassified Orthomyxoviridae.  
OX NCBI\_TaxID=11309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/PHILIPPINES/2/82/BS/ML10 (H3N2);  
RA MEDLINE-97300854; PubMed-9155874;  
RA Hartley C.A., Reading P.C., Ward A.C., Anders E.M.;  
RT "Changes in the hemagglutinin molecule of influenza type A (H3N2)  
RT virus associated with increased virulence for mice."  
RL Arch. Virol. 142:75-88(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/PHILIPPINES/2/82/BS/ML10 (H3N2);  
RA MEDLINE-97456249; PubMed-9311563;  
RA Ward A.C.;  
RT "Virulence of Influenza A virus for mouse lung."  
RL Virus Genes 14:187-194(1997).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL: U08905; AAC79579.1; -  
DR HSSP; P03437; 2YIU.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin.1.  
DR PRINTS: PR00329; HEMAGGLUTN12.  
DR ProDom: PD000225; Hemagglutn.1.  
KW Envelope protein; Glycoprotein; Hemagglutinin.  
FT NON-TER 1  
FT CHAIN 1  
FT CHAIN 1  
FT CHAIN 1  
SQ SEQUENCE 550 AA; 61745 MW; 692A49DE678AC4BC CRC64;

Query Match 92.3%; Score 132; DB 12; Length 550;  
Best Local Similarity 95.7%; Pred. No. 4e-10;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFENGWEGMIDGWG 24  
DB 330 GLFGAIGFENGWEGMIDGWG 352

RESULT 11  
ID 082498 PRELIMINARY; PRT; 550 AA.  
AC 082498;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE HEMAGGLUTININS HA1 AND HA2 (FRAGMENT).  
OS Influenza A virus.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCBI\_TaxID=11320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/PHILIPPINES/2/82 (H3N2);  
RA Hartley C.A., Ward A.C., Anders E.M.;  
RT "Virulence of influenza virus for mice is associated with loss of  
RT oligosaccharide from the hemagglutinin molecule."  
RL Submitted (Apr-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A/PHILIPPINES/2/82 (H3N2);

RA MEDLINE=88185444; PubMed=3356226;  
RT Nakajima S., Takeuchi Y., Nakajima K.;  
RT "Location on the evolutionary tree of influenza H3 haemagglutinin  
genes of Japanese strains isolated during the 1985-6 season.";  
RL Epidemiol. Infect. 100:301-310(1988).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL: U08858; AAA18781.1; -.  
DR HSP: P03437; 2V1U.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin.1.  
DR PRINTS: PR00329; HEMAGGLUTIN1.  
DR Prodom: PD000225; Hemagglutn.1.  
KM Envelope protein; Glycoprotein; Hemagglutinin.  
FT NON\_TER 1 1  
FT CHAIN 1 328 HA1.  
FT CHAIN 330 550 HA2.  
SQ SEQUENCE 550 AA; 61802 MW; 114413B1CE5A1F6A CRC64;

Query Match 92.3%; Score 132; DB 12; Length 550;  
Best Local Similarity 95.7%; Pred. No. 4e-10;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWS 24  
1:|||||  
DB 330 GIFGAIAGFIENGWEGMIDGWS 352

RESULT 12  
Q82496 PRELIMINARY; PRT; 566 AA.  
AC Q82496;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HEMAGGLUTININ.  
GN HA.  
OS Influenza A virus.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCBI\_TaxID=11320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/SWINE/ANGE-GARDIEN/150/90(H3N2);  
RX MEDLINE=95205091; PubMed=7897358;  
RA Bikour M.H., Frost E.H., Deslandes S., Talbot B., Weber J.M.,  
RA Elazhary Y.;  
RT "Recent H3N2 swine influenza virus with haemagglutinin and  
nucleoprotein genes similar to 1975 human strains.";  
RL J. Gen. Virol. 76:697-703(1995).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL: U07145; AAA85781.1; -.  
DR HSP: P03437; 2V1U.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin.1.  
DR PRINTS: PR00329; HEMAGGLUTIN1.  
DR Prodom: PD000225; Hemagglutn.1.  
KM Envelope protein; Glycoprotein; Hemagglutinin.  
SQ SEQUENCE 566 AA; 63362 MW; 1F8B319A567E2FEF CRC64;

Query Match 92.3%; Score 132; DB 12; Length 566;  
Best Local Similarity 95.7%; Pred. No. 4.1e-10;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWS 24  
1:|||||  
DB 346 GIFGAIAGFIENGWEGMIDGWS 368

RESULT 13  
Q03909 PRELIMINARY; PRT; 571 AA.  
AC Q03909;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HEMAGGLUTININ PRECURSOR.  
GN HA.  
OS Influenza A virus.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCBI\_TaxID=11320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Guo Y., Wang M., Kawacha Y., Gorman O.T., Ito T., Webster R.G.;  
RL Submitted (xxx-1991) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION.  
DR EMBL: M65018; AAA43151.1; -.  
DR HSP: P03437; 2V1U.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin.1.  
DR PRINTS: PR00329; HEMAGGLUTIN1.  
DR Prodom: PD000225; Hemagglutn.1.  
KM Envelope protein; Hemagglutinin; Glycoprotein; Signal.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 349 HA1 CHAIN.  
FT CHAIN 350 571 HA2 CHAIN.  
FT DISULFID 36 487 BY SIMILARITY.  
FT DISULFID 73 298 BY SIMILARITY.  
FT DISULFID 85 97 BY SIMILARITY.  
FT DISULFID 160 494 BY SIMILARITY.  
SQ SEQUENCE 571 AA; 64104 MW; 718DA0F291CE349 CRC64;

Query Match 92.3%; Score 132; DB 12; Length 571;  
Best Local Similarity 95.7%; Pred. No. 4.1e-10;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWS 24  
1:|||||  
DB 351 GIFGAIAGFIENGWEGMIDGWS 373

RESULT 14  
Q67050 PRELIMINARY; PRT; 109 AA.  
AC Q67050;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HEMAGGLUTININ (FRAGMENT).  
GN HA.  
OS Influenza A virus.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCBI\_TaxID=11320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/FUKUOKA/C29/85 (H3N2);  
RX MEDLINE=81030852; PubMed=7421990;  
RA Getling M.-J., Bye J., Skene J., Waterfield M.;  
RT "Cloning and dna sequence of double-stranded copies of haemagglutinin  
genes from h2 and h3 strains elucidates antigenic shift and drift in  
human influenza virus.";  
RL Nature 287:301-306(1980).  
RN [2]



RP SEQUENCE FROM N.A.  
 RC STRAIN-A/FUKUOKA/C29/85 (H3N2);  
 RX MEDLINE=93233219; PubMed=7682624;  
 RA Okuno Y., Isegawa Y., Sasao F., Ueda S.;  
 RT "A common neutralizing epitope conserved between the hemagglutinins of  
 influenza A virus H1 and H2 strains."  
 RL J. Virol. 67:2552-2558(1993).  
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
 DR EMBL: D13581; BAA02776.1; -.  
 DR HSSP: P03437; 1H7M.  
 DR InterPro: IPR001364; Hemagglutn.  
 DR Pfam: PF00509; Hemagglutinin.1.  
 DR PRINTS: PR00329; HEMAGGLUTN12.  
 DR ProDom: PD000225; Hemagglutn.1.  
 KW Envelope protein; glycoprotein; Hemagglutinin.  
 FT NON\_TER 1  
 FT SEQUENCE 109 AA; 12305 MW; 17EC66753C48672F CRC64;

Query Match 91.6%; Score 131; DB 12; Length 109;  
 Best Local Similarity 91.3%; Pred. No. 8.9e-11;  
 Matches 21; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIENGWEGMDGNYG 24  
 DB 40 GIFGAIAGFIENGWEGMDGNYG 62

RESULT 15  
 ID 067051 PRELIMINARY; PRT; 109 AA.  
 AC O67051;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE HEMAGGLUTININ (FRAGMENT).  
 GN HA.  
 OS Influenza A virus.  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza virus A and B group; Influenza A viruses.  
 OX NCBI\_TaxID=11320;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A/SICHUAN/2/87 (H3N2);  
 RX MEDLINE=81030852; PubMed=7421990;  
 RA Gething M.-J., Bye J., Skehel J., Waterfield M.;  
 RT "Cloning and dna sequence of double-stranded copies of haemagglutinin  
 genes from h2 and h3 strains elucidates antigenic shift and drift in  
 human influenza virus."  
 RL Nature 287:301-306(1980).  
 RN 12  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A/SICHUAN/2/87 (H3N2);  
 RX MEDLINE=93233219; PubMed=7682624;  
 RA Okuno Y., Isegawa Y., Sasao F., Ueda S.;  
 RT "A common neutralizing epitope conserved between the hemagglutinins of  
 influenza A virus H1 and H2 strains."  
 RL J. Virol. 67:2552-2558(1993).  
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
 DR EMBL: D13582; BAA02777.1; -.  
 DR HSSP: P03437; 1H9E.  
 DR InterPro: IPR001364; Hemagglutn.  
 DR Pfam: PF00509; Hemagglutinin.1.  
 DR PRINTS: PR00329; HEMAGGLUTN12.

DR ProDom: PD000225; Hemagglutn.1.  
 KW Envelope protein; glycoprotein; Hemagglutinin.  
 FT NON\_TER 1  
 FT SEQUENCE 109 AA; 12293 MW; 17EC66752DB8672F CRC64;

Query Match 91.6%; Score 131; DB 12; Length 109;  
 Best Local Similarity 91.3%; Pred. No. 8.9e-11;  
 Matches 21; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIENGWEGMDGNYG 24  
 DB 40 GIFGAIAGFIENGWEGMDGNYG 62

Search completed: July 1, 2002, 06:30:58  
 Job time: 690 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:19:23 ; Search time 98.97 seconds  
(without alignments)  
26.935 Million cell updates/sec

Title: US-09-461-684-3  
Perfect score: 143  
Sequence: 1 GGLFGAIGAFIENGWEGKIDGWYG 24

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	143	100.0	24	21	AA13782
2	143	100.0	26	18	AA26004
3	143	100.0	26	18	AAW34269
4	143	100.0	26	22	AA870090
5	143	100.0	44	21	AA13784
6	134	93.7	23	18	AAW29226
7	134	93.7	25	17	AAW98016
8	134	93.7	26	13	AA26005
9	134	93.7	26	18	AAW34270
10	134	93.7	26	22	AA870091
11	134	93.7	28	17	AAW98024

12	134	93.7	30	21	AAW98680	Core polypeptide f
13	134	93.7	30	22	ABR01088	Viral DP178/107-11
14	134	93.7	30	22	ABR02537	Viral core polypep
15	134	93.7	30	22	AAU13634	DP178-11like/DP107-1
16	134	93.7	30	22	AAW78081	Core polypeptide T
17	134	93.7	175	15	AAW5797	Sequence of BNA2.
18	134	93.7	223	22	AAU14032	Peptide sequence f
19	134	93.7	347	15	AAW63591	Stem region of A2/
20	134	93.7	566	15	AAW63590	Full length H3N2 i
21	134	93.7	685	22	AAW07893	Modified clostridi
22	132	92.3	29	20	AAW29710	Influenza virus an
23	132	92.3	221	14	AAW38864	Sequence of the HA
24	132	92.3	221	14	AAW38865	Sequence of the HA
25	132	92.3	221	15	AAW60230	Influenza haemaggl
26	132	92.3	221	15	AAW60221	Influenza haemaggl
27	132	92.3	306	14	AAW38867	Sequence of fusion
28	132	92.3	306	15	AAW60194	Immunogenic fragme
29	132	92.3	386	5	AAW40615	Sequence of the X-
30	131	91.6	565	8	AAW70711	Equine influenza v
31	131	91.6	565	11	AAW04943	Equine hemagglutin
32	131	91.6	565	19	AAW44946	EIV Fontainebleau
33	131	91.6	565	21	AAW70056	Wild type equine i
34	131	91.6	565	21	AAW70057	Cold-adapted equin
35	131	91.6	566	19	AAW68406	SIV strain H3N2 ha
36	131	91.6	570	8	AAW70710	Equine influenza v
37	131	91.6	570	11	AAW04940	Equine hemagglutin
38	131	91.6	570	18	AAW01669	Influenza A/Beljin
39	131	91.6	570	20	AAW75441	Influenza virus A/
40	131	91.6	570	22	AAW04951	Influenza virus A/
41	131	91.6	571	18	AAW01673	Influenza A/Shanno
42	131	91.6	571	18	AAW01676	Influenza A/Shanno
43	131	91.6	571	20	AAW5445	Influenza virus A/
44	131	91.6	571	20	AAW5448	Influenza virus A/
45	131	91.6	571	22	AAW04955	Influenza virus A/

## ALIGNMENTS

RESULT 1  
AAB13782 standard; peptide: 24 AA.  
XX AAB13782;  
AC AAB13782;  
XX 10-NOV-2000 (first entry)  
DT Soluble peptide antigen HA.  
XX  
DE  
XX  
XX HA peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;  
KW major histocompatibility complex class I; MHC class I; antigen; tumour;  
KW prostate; breast; multiple myeloma.  
XX  
XX Unidentified.  
OS  
XX  
XX WO200035949-A1.  
PN  
XX  
PD 22-JUN-2000.  
XX  
PF 14-DEC-1999; 99WO-US29724.  
XX  
XX 14-DEC-1996; 98US-0112324.  
PR (DEND-) DENDREON CORP.  
XX  
XX Laus R, Hakim I, Vidovic D;  
XX WPI: 2000-442365/38.  
XX  
XX Antigen modified by the covalent addition of a peptide that  
XX facilitates entry into antigen presenting cells, useful for producing  
XX compositions for immunizing against tumors and pathogens -

PS Claim 2; Page 26; 34pp; English.

XX The present invention relates to compositions of modified soluble protein  
CC antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)  
CC response i.e. a major histocompatibility complex (MHC) class I molecule  
CC response. The protein antigen is modified by the covalent addition of a  
CC peptide sequence which facilitate entry of the antigen into antigen  
CC presenting cells (APCs). The present sequence is one such peptide  
CC sequence which can be used to modify the soluble antigens. The present  
CC sequence is peptide HA. The modified antigen composition may be used for  
CC immunising against, or treating a tumour e.g. prostate and breast  
CC carcinoma or multiple myeloma, or pathogen in mammals.

XX Sequence 24 AA;

Query Match 100.0%; Score 143; DB 21; Length 24;  
Best Local Similarity 100.0%; Pred. No. 1.2e-14;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGLFGAIAFTENGEGMIDGWYG 24  
Db 1 cglfgaiaftengegmidgwyg 24

RESULT 2

AAW34269 ID AAW34269 standard; peptide; 26 AA.

XX AAW34269;

XX 26-MAY-1998 (first entry)

XX Influenza fusion peptide #1.

XX Haemagglutinin; cystein; liposome.

XX Synthetic.

XX EP497997-A.

XX 12-AUG-1992.

XX 02-FEB-1991; 91EP-0101414.

XX 02-FEB-1991; 91EP-0101414.

XX (NIKA-) NIKA HEALTH PROD LTD.

XX Gluck R, Herrmann P, Klein P;

XX WPI; 1992-270078/33.

XX Dryer-conc. phospholipid bilayer vesicle with cell-specific

XX markers on membrane - where markers have at least 90 per cent

XX biological activity, used as pharmaceuticals against e.g. cancer

XX and AIDS

XX Disclosure; Fig 2; 13pp; English.

XX The sequences given in AAW34269-19 are influenza fusion peptides,  
CC derived from the influenza virus haemagglutinin gene which were  
CC used for the preparation of synthetic membrane vesicles. The  
CC arrangement of at least one, pref. three cystein residues at one  
CC for the fusion of the liposome to the target cell membrane. The  
CC liposomes produced using these peptides can contain at least one  
CC active drug and can be used to target cells infected with cancer  
CC or AIDS.

XX Sequence 26 AA;

Query Match 100.0%; Score 143; DB 13; Length 26;  
Best Local Similarity 100.0%; Pred. No. 1.4e-14;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGLFGAIAFTENGEGMIDGWYG 24  
Db 3 cglfgaiaftengegmidgwyg 26

RESULT 3

AAW34269 ID AAW34269 standard; peptide; 26 AA.

XX AAW34269;

XX 14-MAY-1998 (first entry)

XX Synthetic lipid vesicle fusion peptide 1.

XX Virosome; vesicle; cationic lipid bilayer; viral fusion peptide;

XX drug delivery system; membrane; gene therapy; diagnosis; treatment;

XX cancer; leukaemia; viral infection.

XX Synthetic.

XX WO9741834-A1.

XX 13-NOV-1997.

XX 04-MAY-1997; 97WO-EP02268.

XX 08-MAY-1996; 96EP-0107282.

XX (NIKA-) NIKA HEALTH PROD LTD.

XX Glueck R, Klein P, Maeltl ER;

XX WPI; 1997-558673/51.

XX Vesicle with cationic lipid bilayer that includes viral fusion

XX peptide - used for delivery of genetic material to cells, especially

XX for gene therapy of cancer, leukaemia and viral infections

XX Disclosure; Page 10; 52pp; English.

XX Peptides AAW34269-W34284 represent novel lipid vesicles with positively  
CC charged lipid bilayer membranes composed of a cationic and/or  
CC polycationic lipid and at least one natural or synthetic viral fusion  
CC peptide integrated in, or covalently linked to, the membrane. Such  
CC peptides are used as drug delivery systems, preferably for  
CC (non-)specific delivery of genetic material to target cells or tissues,  
CC particularly for diagnosis, treatment (especially antisense treatment)  
CC of cancer, leukaemia and viral infections in humans or animals. Genetic  
CC material is delivered, without infection, to resting or proliferating  
CC cells, in vitro or in vivo. When the genetic material is an antisense  
CC molecule, it is targeted to mRNA encoding a (proto)oncogene. The  
CC continuous lipid layer does not leak. The peptides do not need to fuse  
CC with, or destabilise, plasma membranes in order to enter the cytoplasm,  
CC since the fusion peptide ensures cell penetration by endocytosis (after  
CC which fusion of the vesicle and endosomal membrane occurs). The genetic  
CC material thus has a greater chance of reaching the nucleus before it is  
CC degraded or expelled. Transfer of the material is 1000-20000 times more  
CC efficient than when using liposomes or conventional virosomes, so  
CC smaller doses can be used, avoiding possible toxicity associated with the  
CC genetic material.

XX Sequence 26 AA;

Query Match 100.0%; Score 143; DB 18; Length 26;  
Best Local Similarity 100.0%; Pred. No. 1.4e-14;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGLFGAIGFIENGWEGMIDGWY 24  
Db 3 cglfgaigfiengwewgmldgwy 26

RESULT 4  
AAB70090  
ID AAB70090 standard; peptide: 26 AA.  
XX  
XX AAB70090;  
DT 14-MAY-2001 (first entry)  
DE Cationic virosome crosslinker #1 for polypeptide binding.  
XX  
XX Cationic virosome; cytosolic; gene therapy; lipid bilayer vesicle;  
KW viral haemagglutinin; drug delivery; cancer; leukaemia; viral infection;  
KM crosslinker.  
XX  
XX Unidentified.  
XX  
XX N2504444-A.  
XX  
XX 24-NOV-2000.  
XX  
XX 10-MAY-2000; 2000NZ-0504444.  
XX  
XX 10-MAY-2000; 2000NZ-0332666.  
XX  
XX (NIKA-) NIKA HEALTH PROD LTD.  
XX  
XX Walth ER, Gluck R, Klein P;  
DR WPI: 2001-233042/24.  
XX  
XX Lipid bilayer vesicle, useful for delivering drugs to target cells such  
PT as cancer, leukemic, or virally infected cells, comprises viral  
PT glycoproteins in positively charged membrane  
XX  
XX PS Disclosure: Page 14; 41pp; English.

CC The present sequence is a crosslinker which is capable of linking to the  
CC surface of a novel cationic virosome and is capable of binding  
CC polypeptides. The cationic virosome is a lipid bilayer vesicle comprising  
CC a membrane with a net positive charge and contains 5-30 weight % based on  
CC total lipid, 1,3-dioleoyloxy-2-(6-carboxy-spermyl)-propyl-amine  
CC (DOSPER), together with other lipids and at least one active fusogenic  
CC peptide. The fusogenic peptide is a viral haemagglutinin that causes the  
CC vesicle to be internalised by target cells through phagocytosis or  
CC endocytosis. The virosome is useful for delivering a desired drug or  
CC substance, preferably a nucleic acid, to target cells (resting or  
CC proliferating mammalian cells) such as cancer cells, leukemic cells or  
CC virally infected cells in vitro, in diagnostic or medical applications  
CC and for the manufacture of medicament for carrying out non-infectious  
CC gene therapy.  
XX  
XX SQ Sequence 26 AA;

Query Match 100.0%; Score 143; DB 22; Length 26;  
Best Local Similarity 100.0%; Pred. No. 1.4e-14;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGLFGAIGFIENGWEGMIDGWY 24  
Db 3 cglfgaigfiengwewgmldgwy 26

RESULT 5  
AAB13784  
ID AAB13784 standard; peptide: 44 AA.  
XX  
XX AAB13784;

XX 10-NOV-2000 (first entry)  
DT  
XX Soluble tandem HA/ PK peptide conjugate.  
DE  
XX PK peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;  
KW major histocompatibility complex class I; MHC class I; antigen; tumour;  
KM prostate; breast; multiple myeloma; HA peptide.  
XX  
XX Unidentified.  
XX  
XX WO200035949-A1.  
XX  
XX 22-JUN-2000.  
XX  
XX 14-DEC-1999; 99WO-US29724.  
XX  
XX 14-DEC-1998; 98US-0112324.  
XX  
XX (DEND-) DENDREON CORP.  
XX  
XX Laus R, Hakim I, Vidovic D;  
XX  
XX WPI: 2000-442365/38.  
XX  
XX Antigen modified by the covalent addition of a peptide that  
PT facilitates entry into antigen presenting cells, useful for producing  
PT compositions for immunizing against tumors and pathogens  
XX  
XX PS Claim 2; Page 26; 34pp; English.

CC The present invention relates to compositions of modified soluble protein  
CC antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)  
CC response i.e. a major histocompatibility complex (MHC) class I molecule  
CC response. The protein antigen is modified by the covalent addition of a  
CC peptide sequence which facilitate entry of the antigen into antigen  
CC presenting cells (APCs). The present sequence is one such peptide  
CC sequence which can be used to modify the soluble antigens. The present  
CC composition is tandem HA/ PK peptide conjugate. The modified antigen  
CC composition may be used for immunising against, or treating a tumour e.g.  
CC prostate and breast carcinoma or multiple myeloma, or pathogen in  
CC mammals.  
XX  
XX SQ Sequence 44 AA;

Query Match 100.0%; Score 143; DB 21; Length 44;  
Best Local Similarity 100.0%; Pred. No. 2.5e-14;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGLFGAIGFIENGWEGMIDGWY 24  
Db 1 cglfgaigfiengwewgmldgwy 24

RESULT 6  
AAW29226  
ID AAW29226 standard; peptide: 23 AA.  
XX  
XX AAW29226;  
XX  
XX 02-MAR-1998 (first entry)  
XX  
XX Membrane active acidic peptide INF6 from WO9730170.  
DE  
XX Membrane active; acidic peptide; cationic lipid; transfection;  
KW gene therapy; tumour; vaccine; cytokine gene.  
XX  
XX Synthetic.  
OS  
XX Influenza virus.  
XX  
XX WO9730170-A1.

PD 21-AUG-1997.  
XX  
XX 13-FEB-1997; 97WO-EP00649.  
PF  
XX 15-FEB-1996; 96DE-4005548.  
PR  
XX (BOEH ) BOEHRINGER INGELHEIM INT GMBH.  
PA  
XX Kichler A, Mechtler K, Wagner E;  
PI  
XX WPI; 1997-425044/39.  
DR  
XX Composition containing nucleic acid complexed with cationic lipid  
PT and a membrane-active, acidic peptide - used to increase efficiency  
PT of transfection by destabilising the endosomal membrane, especially  
PT useful for gene therapy  
XX  
XX Claim 4; Page 33; 63pp; German.  
PS  
XX A new composition has been developed for transfecting higher eukaryotic  
CC cells contains: (i) a complex of nucleic acid to be expressed plus a  
CC suboptimal concentration of one or more cationic lipids, and optionally  
CC at least one auxiliary lipid; and (ii) the new feature, at least one  
CC membrane-active, acidic peptide; where the ratio total positive  
CC charges:total-negative charges in the composition is about 0-3,  
CC preferably 0-2. The present sequence represents a specifically claimed  
CC peptide for use in the composition. The composition is used for  
CC transfection in vitro, in vivo or ex vivo, particularly for gene  
CC therapy. A typical application is production of tumor vaccines from  
CC autologous cells transfected with a cytokine gene. The acidic peptide  
CC destabilises the endosomal membrane and significantly increases  
CC transfection efficiency, e.g. by as much as 1000 times, equivalent to  
CC that achieved with the optimal amount of cationic lipid. The acidic  
CC peptide also makes the transfection complex less sensitive to serum.  
CC  
XX Sequence 23 AA:  
SQ  
Query Match 93.7%; Score 134; DB 18; Length 23;  
Best Local Similarity 100.0%; Pred. No. 2.6e-13;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2 GLEGAIAGFIEGMEGMDWG 24  
DB 1 glfgaiafienegwgmldwyg 23  
RESULT 7  
AAR98016  
ID AAR98016 standard; peptide: 25 AA.  
XX  
XX AAR98016;  
AC  
XX 12-NOV-1996 (first entry)  
DT  
XX Fusedogenic peptide derived from HA-2 spike glycoprotein.  
DE  
XX Fusedogenic peptide; spike glycoprotein; enveloped virus; vaccine; EMPC;  
XX endosome membrane disruption promoting component; cationic polyamine;  
XX multifunctional molecular complex; foodstuff; herbicide; insecticide;  
XX plant growth regulator; miticide; rodenticide; fungicide; parasiticide;  
XX nematocide; immunisation; pathogen; therapy; autoimmune disease;  
XX hyperproliferating disease.  
XX  
XX Synthetic.  
OS  
XX WO9610038-A1.  
PN  
XX 04-APR-1996.  
PD  
XX 28-SEP-1995; 95WO-US12502.  
PF  
XX 28-SEP-1994; 94US-0314060.  
PR

XX  
XX (APOL-) APOLLON INC.  
PA  
XX Boutin RH;  
PI  
XX WPI; 1996-200887/20.  
DR  
XX New complexes for nucleic acid transfer to target cells - comprising  
PT a nucleic acid compn. and a cationic polyamine with an endosome  
PT membrane disruption component.  
PT  
XX Disclosure; Page 30; 138pp; English.  
PS  
XX AAR98010-R98041 represent fusedogenic peptides derived from spike  
CC glycoproteins of enveloped viruses. These fusedogenic peptides act as  
CC endosome membrane disruption promoting components (EMDPCs). These  
CC sequences were used in a transfer moiety which makes up part of the  
CC multifunctional molecular complex of the invention. The multifunctional  
CC molecular complex is used for the transfer of a nucleic acid composition  
CC to a target cell and comprises the nucleic acid composition and a  
CC transfer moiety comprising one or more cationic polyamine components  
CC bound to the nucleic acid composition. There are several different  
CC cationic polyamines that can be used in the complex, one of which  
CC contains one of these sequences linked via a bridging group to the main  
CC body of the cationic polyamine. The products can be used for delivering  
CC therapeutic agents, vaccines, foodstuffs, herbicides, plant growth  
CC regulators, insecticides, miticides, rodenticides, fungicides,  
CC parasiticides or nematocides. They can also be used for immunising an  
CC individual against a pathogen or for treating an autoimmune disease or  
CC hyperproliferating disease. The complexes provide for a high level of  
CC transfection and expression of the nucleic acid molecules in target  
CC cells.  
XX  
XX Sequence 25 AA:  
SQ  
Query Match 93.7%; Score 134; DB 17; Length 25;  
Best Local Similarity 100.0%; Pred. No. 2.9e-13;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2 GLEGAIAGFIEGMEGMDWG 24  
DB 1 glfgaiafienegwgmldwyg 23  
RESULT 8  
AAR26005  
ID AAR26005 standard; peptide: 26 AA.  
XX  
XX AAR26005;  
AC  
XX 26-JAN-1993 (first entry)  
DT  
XX Influenza fusion peptide #2.  
DE  
XX Haemagglutinin; cysteine; liposome.  
XX  
XX Synthetic.  
OS  
XX BP497997-A.  
XX  
XX 12-AUG-1992.  
PD  
XX 02-FEB-1991; 91EP-0101414.  
PF  
XX 02-FEB-1991; 91EP-0101414.  
PR  
XX (NIKA-) NIKA HEALTH PROD LTD.  
PA  
XX Gluck R, Herrmann P, Klein P;  
PI  
XX WPI; 1992-270078/33.  
DR  
XX

PT Drug-contg. phospholipid bilayer vesicle with cell-specific  
PT markers on membrane - where markers have at least 90 per cent  
PT biological activity, used as pharmaceuticals against e.g. cancer  
PT and AIDS  
PS Disclosure; Fig 2; 13pp; English.  
XX The sequences given in AAR26004-19 are influenza fusion peptides,  
CC derived from the influenza virus haemagglutinin gene which were  
CC used for the preparation of synthetic membrane vesicles. The  
CC arrangement of at least one, pref. three cysteine residues at one  
CC end of these peptides has been found useful for the fusion activity,  
CC for the fusion of the liposome to the target cell membrane. The  
CC liposomes produced using these peptides can contain at least one  
CC active drug and can be used to target cells infected with cancer  
CC or AIDS.  
XX  
SQ Sequence 26 AA;  
Query Match 93.7%; Score 134; DB 13; Length 26;  
Best Local Similarity 100.0%; Pred. No. 3e-13;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GLFGAAGFIENGWEGMIDGNG 24  
DB 1 glfgaagfiengwegmidgng 23  
RESULT 9  
AAR34270  
ID AAR34270 standard; peptide; 26 AA.  
XX  
AC AAR34270;  
XX  
DT 14-MAY-1998 (first entry)  
XX  
DE Synthetic lipid vesicle fusion peptide 2.  
XX  
XX Virosome; vesicle; cationic lipid bilayer; viral fusion peptide;  
KW drug delivery system; membrane; gene therapy; diagnosis; treatment;  
KW cancer; leukaemia; viral infection.  
XX  
OS Synthetic.  
XX  
PN W09741834-A1.  
XX  
PD 13-NOV-1997.  
XX  
PF 04-MAY-1997; 97WC-EP02268.  
XX  
PR 08-MAY-1996; 96EP-0107282.  
XX  
PA (NIKA-) NIKA HEALTH PROD LTD.  
XX  
PI Glueck R, Klein P, Waelti ER;  
XX  
DR WPI; 1997-558673/51.  
XX  
XX Vesicle with cationic lipid bilayer that includes viral fusion  
PT peptide - used for delivery of genetic material to cells, especially  
PT for gene therapy of cancer, leukaemia and viral infections  
XX  
PS Disclosure; Page 10; 52pp; English.  
XX  
CC Peptides AAR4269-W34284 represent novel lipid vesicles with positively  
CC charged lipid bilayer membranes composed of a cationic and/or  
CC polycationic lipid and at least one natural or synthetic viral fusion  
CC peptide integrated in, or covalently linked to, the membrane. Such  
CC peptides are used as drug delivery systems, preferably for  
CC (non)-specific delivery of genetic material to target cells or tissues,  
CC particularly for diagnosis, treatment (especially antisense treatment),  
CC of cancer, leukaemia and viral infections in humans or animals. Genetic

CC material is delivered, without infection, to resting or proliferating  
CC cells, in vitro or in vivo. When the genetic material is an antisense  
CC molecule, it is targeted to mRNA encoding a (proto)oncogene. The  
CC continuous lipid layer does not leak. The peptides do not need to fuse  
CC with, or destabilise, plasma membranes in order to enter the cytoplasm,  
CC since the fusion peptide ensures cell penetration by endocytosis (after  
CC which fusion of the vesicle and endosomal membrane occurs). The genetic  
CC material thus has a greater chance of reaching the nucleus before it is  
CC degraded or expelled. Transfer of the material is 1000-20000 times more  
CC efficient than when using liposomes or conventional virosomes, so  
CC smaller doses can be used, avoiding possible toxicity associated with the  
CC genetic material.  
XX  
SQ Sequence 26 AA;  
Query Match 93.7%; Score 134; DB 18; Length 26;  
Best Local Similarity 100.0%; Pred. No. 3e-13;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GLFGAAGFIENGWEGMIDGNG 24  
DB 1 glfgaagfiengwegmidgng 23  
RESULT 10  
AAB70091  
ID AAB70091 standard; peptide; 26 AA.  
XX  
AC AAB70091;  
XX  
DT 14-MAY-2001 (first entry)  
XX  
DE Cationic virosome crosslinker #2 for polypeptide binding.  
XX  
XX Cationic virosome; cytostatic; gene therapy; lipid bilayer vesicle;  
KW viral haemagglutinin; drug delivery; cancer; leukaemia; viral infection;  
KW crosslinker.  
XX  
XX Unidentified.  
XX  
PN NZ50444-A.  
XX  
PD 24-NOV-2000.  
XX  
PF 10-MAY-2000; 2000NZ-0504444.  
XX  
PR 10-MAY-2000; 2000NZ-0332666.  
XX  
PA (NIKA-) NIKA HEALTH PROD LTD.  
XX  
PI Waeli ER, Glueck R, Klein P;  
XX  
DR WPI; 2001-233042/24.  
XX  
XX Lipid bilayer vesicle, useful for delivering drugs to target cells such  
PT as cancer, leukemic, or virally infected cells, comprises viral  
PT glycoproteins in positively charged membrane -  
XX  
PS Disclosure; Page 14; 41pp; English.  
XX  
XX The present sequence is a crosslinker which is capable of linking to the  
CC surface of a novel cationic virosome and is capable of binding  
CC polypeptides. The cationic virosome is a lipid bilayer vesicle comprising  
CC a membrane with a net positive charge and contains 5-30 weight % based on  
CC total lipids, 1,3-dioleoyloxy-2-(6-carboxy-spermyl)-propyl-amide  
CC (DOSER), together with other lipids and at least one active fusogenic  
CC peptide. The fusogenic peptide is a viral haemagglutinin that causes the  
CC vesicle to be internalised by target cells through phagocytosis or  
CC endocytosis. The virosome is useful for delivering a desired drug or  
CC substance, preferably a nucleic acid, to target cells (resting or  
CC proliferating mammalian cells) such as cancer cells, leukaemic cells or  
CC virally infected cells in vitro, in diagnostic or medical applications







OS Synthetic.

Key

ET Modified-site

Location/Qualifiers  
1

FT	Mod
----	-----

```

30 /note= "N-terminal is substituted by Ac"
    /note= "C-terminal amide"

```

PN WO200151673-A2.

PD 19-JUL-2001.

05-JUL-2000; 2000WO-US35727.

PR 09-JUL-1999; 99US-0350841.

PA (TRIM-) TRIMERIS INC.

PI Jeffs P, Lackey JW, Erickson JB, Lawless MK, Merutka G;  
xx

WPI; 2001-442157/47.

PT identifying a compound that inhibits the formation of or disrupts a  
PT Dp107/Dp178 complex, especially compounds with antileukogenic, antiviral  
PT or intracellular modulatory activity, by detecting the formation of a  
PT Dp107/Dp178 complex -

PS Disclosure; Page 74; 259pp; English.

The present invention relates to peptides which exhibit anti-retroviral activity. The peptides of the invention (AAU12559-AAU14003) comprise Dp178-1-like and Dp107-1-like peptides. The Dp178 peptide corresponds to amino acids 639-673 of the transmembrane protein gp41 from human immunodeficiency virus 1 (HIV-1) isolate LAI. The Dp107 peptide corresponds to amino acids 558-595 of gp41 from HIV-1-LAI. The invention also relates to a method of identifying compounds that inhibit the formation of or disrupts a Dp107/Dp178 complex. The method comprises detecting the formation of a Dp107/Dp178 complex, both in the presence or absence of a test compound, in a reaction mixture containing Dp107 and Dp178 peptides. The method is useful for identifying compounds, including small molecule compounds, which may themselves exhibit antiretroviral, antiviral or intracellular modulatory activity. The Dp178-1-like/Dp107-1-like peptides are useful to inhibit human and non-human retroviral, particularly HIV, transmission to uninfected cells. The present sequence represents one of the Dp178-1-like/Dp107-1-like peptides of the invention.

50 Sequence 30 AA;

Query Match	93.78;	Score 134;	DB 22;	Length 30;
-------------	--------	------------	--------	------------

Best Local Similarity 100.0%; Pred. No. 3.6e-13;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 2 GLFGAAGFTIENGWEGMIDGWYG 24

Db 1 glfgalagfliengwegmidgwyg 23

Search completed: July 1, 2002, 06:19:24  
Job time: 511 sec

Job time: 511 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 1, 2002, 06:28:11 ; Search time 49.45 Seconds

(without alignments)  
87.442 Million cell updates/sec

Title: US-09-461-684-4  
Perfect score: 209  
Sequence: 1 CEAATAAEEAATAAATAAEEA.....KKKKKKKKKKKKKKKKKK 45

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	103	49.3	248	1	HSUR1P
2	102	48.8	517	2	T49173
3	101.5	48.6	442	2	T39683
4	100	47.8	215	2	I52523
5	100	47.8	380	2	T46395
6	98	46.9	206	2	S09388
7	96	45.9	433	2	S25194
8	95	45.5	166	2	T18513
9	91	43.5	229	2	UC7219
10	90.5	43.3	383	2	AB6315
11	90	43.1	153	2	S59591
12	90	43.1	483	2	F71619
13	90	43.1	529	2	T50609
14	89	42.6	208	2	T23778
15	89	42.6	265	2	S19113
16	88	42.1	392	2	T15755
17	88	42.1	107	2	T18726
18	87	41.6	109	2	C86477
19	87	41.6	565	2	T47775
20	86	41.1	347	2	E83525
21	86	41.1	455	2	AB7913
22	85.5	40.9	1280	2	G96796
23	85	40.7	102	2	D64363
24	85	40.7	153	2	S59587
25	85	40.7	241	2	SN0748
26	85	40.7	560	2	T06377
27	85	40.7	849	2	T01285
28	84.5	40.4	228	2	T36379
29	84	40.2	409	2	T24543

30	84	40.2	425	2	T18723	hypothetical prote
31	84	40.2	441	2	A48455	acidic phosphoprot
32	84	40.2	1560	2	T42727	proliferation pore
33	84	40.2	1701	2	T09127	probable erythrocy
34	83	39.7	107	2	A75031	1su ribosomal prote
35	83	39.7	146	2	A13340	hypothetical prote
36	83	39.7	320	2	S61586	probable membrane
37	83	39.7	421	2	JV0057	tola protein - Esc
38	83	39.7	625	2	T39019	probable mitochond
39	83	39.7	153	2	S59125	histone H2B (valid
40	82	39.2	474	2	T38485	centromere/microtu
41	81.5	39.0	628	2	T08942	proton pump intera
42	81	38.8	111	2	F71216	probable ribosomal
43	81	38.8	218	2	UC7220	nuclear protein SR
44	80.5	38.5	1002	2	S70292	FUN12 protein - ye
45	80	38.3	111	2	E72524	probable ribosomal

## ALIGNMENTS

RESULT 1  
HSUR1P  
histone H1, gonadal - sea urchin (Parechinus angulosus)  
C:Species: Parechinus angulosus (angulate urchin)  
C:Date: 31-Mar-1980 #sequence\_revision 31-Mar-1980 #text\_change 16-Feb-1997  
C:Accession: A91090; A91091; A02566  
E:Strickland, W.N.; Strickland, M.; de Groot, P.C.; von Holt, C.; Wittmann-Liebold, B  
Eur. J. Biochem. 104, 559-566, 1980  
A:Title: The primary structure of histone H1 from sperm of the sea urchin Parechinus  
terminal cyanogen bromide peptides.  
A:Reference number: A91090; M0156831  
A:Contents: sequence of residues 1-84  
A:Accession: A91090  
A:Molecule type: protein  
A:Residues: 1-248 <STR>  
E:Strickland, W.N.; Strickland, M.; Brandt, W.F.; von Holt, C.; Lehmann, A.; Wittmann  
Eur. J. Biochem. 104, 567-578, 1980  
A:Title: The primary structure of histone H1 from sperm of the sea urchin Parechinus  
A:Reference number: A91091; M0156832  
A:Accession: A91091  
A:Molecule type: protein  
A:Residues: 80-248 <STR>  
A:Note: 144-Arg was also found  
C:Superfamily: histone H1  
C:Keywords: DNA binding; nucleosome; sperm

Query Match 49.3%; Score 103; DB 1; Length 248;  
Best Local Similarity 59.5%; Pred. No. 0.087;  
Matches 25; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

QY 2 EAAATAAEEAATAAATAAATAAEEA.....KKKKKKKKKKKKKKKKKK 43  
DB 146 KAAAKRKAALAKRKAATAAKRKAARAKKAKKAKKAKKAKKAKK 187

RESULT 2  
T49173  
hypothetical protein T20N10.250 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 08-Dec-2000  
C:Accession: T49173  
R:D'Angelis, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.  
submitted to the protein Sequence Database, April 2000  
A:Reference number: Z25017  
A:Accession: T49173  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-517 <DNA>  
A:Cross-references: EMBL:AL353032; GSPDB:GN00061; ATSP:T20N10.250  
A:Experimental source: Cultivar Columbia; BAC clone T20N10  
C:Genetics:







Mon Jul 1 13:55:04 2002

us-09-461-684-4.rpr

Page 5

```

Oy      3  AAAAAAEEAAAAAEEAAAAAKKKKKKKK 34
          ||| | ||||| : ||| : | : |
Db      176 AAAAAEAAAAEAAAAAKKRAAAEAKAKKADK 207

```

Search completed: July 1, 2002, 06:28:12  
Job time: 714 sec













KW Bromodomain; Repeat: Nuclear protein.  
 FT DOMAIN 56 115  
 FT DOMAIN 326 338  
 FT DOMAIN 487 535  
 FT DOMAIN 676 725  
 FT CONFLICT 465 466  
 FT SEQUENCE 726 AA; 79541 MW; 64F526FC3C1033AA CRC64;

Query Match 40.2%; Score 84; DB 1; Length 726;  
 Best Local Similarity 54.3%; Pred. No. 2;  
 Matches 19; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 11 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  
 DB 472 AVEGQALASQAPVNPVKKKKKKKKKKKKKK 45

RESULT 11  
 RL12\_PYRAB STANDARD; PRT; 107 AA.  
 AC G9UXS6;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 50S ribosomal protein L12P.  
 GN RPL12P OR PAB1168.  
 OS Pyrococcus abyssi.  
 OC Archaea; Euryarchaeota; Thermococcales; Pyrococcus.  
 OX NCBI\_TaxId=29292;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ORSAY;  
 RA Heilig R.;

RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution."  
 RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: SEEKS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR ACCURATE TRANSLATION (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
 CC EMBL: AJ248288; CAB50687.1; -  
 DR InterPro; IPR001813; 60S\_ribosomal.  
 DR Pfam; PF00428; 60S\_ribosomal; 1.  
 KW Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 107 AA; 11300 MW; CDA23930BAD4ED CRC64;

Query Match 39.7%; Score 83; DB 1; Length 107;  
 Best Local Similarity 44.2%; Pred. No. 0.34;  
 Matches 19; Conservative 12; Mismatches 12; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  
 DB 55 EKAAAPVAAAPAAAPAEAPAEKKEKEKEKEKEEVESE 97

RESULT 12  
 YD33\_YEAST STANDARD; PRT; 320 AA.  
 ID YD33\_YEAST  
 AC Q12117;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical 36.2 kDa protein in RAD28-LYS14 intergenic region.  
 GN YDR033W OR YD9673.03.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxId=4932;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Arnold W., Becker A., Jaeger W., Kuester H., Nussbaumer B.;  
 RN Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 RL (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE ARCHAEL OPSIN FAMILY. HSP30 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
 CC EMBL: Z74329; CA98855.1; -  
 DR EMBL; Z68196; CA92370.1; -  
 DR SGD; S0002440; YDR033W.  
 DR InterPro; IPR01425; Bac\_rhodopsin.  
 DR Pfam; PF01036; Bac\_rhodopsin; 1.  
 KW Hypothetical protein; Transmembrane.  
 FT TRANSMEM 35 55  
 FT TRANSMEM 63 83  
 FT TRANSMEM 117 137  
 FT TRANSMEM 142 162  
 FT TRANSMEM 168 188  
 FT TRANSMEM 205 225  
 FT TRANSMEM 239 259  
 FT DOMAIN 300 318  
 SQ SEQUENCE 320 AA; 36190 MW; 4311F64DA6AA209F CRC64;

Query Match 39.7%; Score 83; DB 1; Length 320;  
 Best Local Similarity 48.8%; Pred. No. 1.3;  
 Matches 21; Conservative 4; Mismatches 14; Indels 4; Gaps 1;

QY 7 AAAAAAAAAAAAAAAAAAA-----AKKKKKKKKKKKKKKKKKKKK 45  
 DB 277 AEAVAPAPAPVSPRPATPMLSKDKKKSKSKSKSKSKSE 319

RESULT 13  
 ID TOLA\_ECOLI STANDARD; PRT; 421 AA.  
 AC P19934;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE TOLA protein.  
 GN TOLA OR CIM OR EXCC OR LKY OR B0739.  
 OS Escherichia coli.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia.  
 OX NCBI\_TaxId=562;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JM105;  
 RX MEDLINE=90078104; PubMed=2687247;  
 RA Levegood S.K., Webster R.E.;  
 RT "Nucleotide sequences of the *tola* and *tolB* genes and localization of their products, components of a multistep translocation system in





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: July 1, 2002, 06:30:58 ; Search time 83.98 Seconds  
(without alignments)  
92.698 Million cell updates/sec

Title: US-09-461-684-4  
Sequence: 1 CCAAAAEAAAAAEAAAAAE.....KKKKKKKKKKKKKKKKKK 45

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP rhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP virus:\*  
16: SP bacteriap:\*  
17: SP archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	118	56.5	2451	5 Q9VG05	Q9VG05 drosophila
2	105	50.2	129	11 Q35807	Q35807 rattus norv
3	105	50.2	720	4 Q9H607	Q9H607 homo sapien
4	102	48.8	467	10 Q9L182	Q9L182 euglena gra
5	102	48.8	517	10 Q9LXR2	Q9LXR2 arabidopsis
6	101.5	48.3	354	3 Q14347	Q14347 schizosach
7	101	48.3	531	6 Q95LV6	Q95LV6 macaca fasc
8	100	47.8	128	3 Q9P529	Q9P529 neurospora
9	100	47.8	168	4 Q9H5V6	Q9H5V6 homo sapien
10	100	47.8	215	11 Q64075	Q64075 rattus sp.
11	100	47.8	260	10 Q9LGT9	Q9LGT9 arabidopsis
12	100	47.8	380	4 Q9NT34	Q9NT34 homo sapien
13	99.5	47.6	377	5 Q9GN12	Q9GN12 leishmania
14	99	47.4	686	4 Q9NXY0	Q9NXY0 homo sapien
15	97	46.4	657	4 Q9HC48	Q9HC48 homo sapien
16	95	45.5	257	4 Q9H5Y3	Q9H5Y3 homo sapien

17	92.5	44.3	372	2 Q9WXX1	Q9WXX1 pseudomonas
18	91	43.5	229	11 Q9JW93	Q9JW93 mus musculus
19	91	43.5	263	4 Q9BU76	Q9BU76 homo sapien
20	90.5	43.3	368	10 Q944I9	Q944I9 arabidopsis
21	90.5	43.3	383	10 Q9LMT3	Q9LMT3 arabidopsis
22	90	43.1	166	5 Q9G281	Q9G281 parecinius
23	90	43.1	180	5 Q25636	Q25636 parecinius
24	90	43.1	483	5 Q96148	Q96148 plasmodium
25	90	43.1	529	4 Q9NFX4	Q9NFX4 homo sapien
26	90	43.1	628	4 Q9H5M5	Q9H5M5 homo sapien
27	90	43.1	749	5 Q967D9	Q967D9 drosophila
28	90	43.1	1062	5 Q960C4	Q960C4 drosophila
29	90	43.1	1430	5 Q9W0B0	Q9W0B0 drosophila
30	90	43.1	1616	4 Q96PH3	Q96PH3 homo sapien
31	89	42.6	83	11 Q9ER82	Q9ER82 mus musculus
32	89	42.6	158	10 Q9MBF7	Q9MBF7 illium long
33	89	42.6	208	5 Q93901	Q93901 caenorhabdi
34	89	42.6	265	10 Q93598	Q93598 chlamydomon
35	88	42.1	260	11 Q991X5	Q991X5 mus musculu
36	88	42.1	392	5 Q18390	Q18390 caenorhabdi
37	88	42.1	409	5 Q903S6	Q903S6 caenorhabdi
38	88	42.1	1166	5 Q9Y114	Q9Y114 drosophila
39	87.5	41.9	533	10 Q9MAB3	Q9MAB3 arabidopsis
40	87	41.6	107	10 Q9L0P6	Q9L0P6 arabidopsis
41	87	41.6	565	10 Q9LD90	Q9LD90 arabidopsis
42	86	41.1	118	13 Q9IAZ3	Q9IAZ3 paratichthy
43	86	41.1	347	2 Q9RKL9	Q9RKL9 streptomyc
44	86	41.1	455	5 Q61747	Q61747 caenorhabdi
45	85.5	40.9	1173	5 Q95XK7	Q95XK7 caenorhabdi

## ALIGNMENTS

RESULT	ID	Q9VG05	PRELIMINARY:	PRT:	2451 AA.
Q9VG05	Q9VG05	Q9VG05	Q9VG05	Q9VG05	Q9VG05
AC	Q9VG05	Q9VG05	Q9VG05	Q9VG05	Q9VG05
DT	01-MAY-2000	(TREMBLrel. 13, Created)			
DT	01-MAR-2001	(TREMBLrel. 16, Last sequence update)			
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)			
DE	CG7518	PROTEIN.			
GN	CG7518				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
OC	Ephydroidea; Drosophilidae; Drosophila.				
OX	NCBI_TaxID=7227;				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-BERKELEY;				
RX	MEDLINE=20196006; PubMed=10731132;				
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,				
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,				
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,				
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,				
RA	Brandon R.C., Rogers Y.H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,				
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mklos G.L.G.,				
RA	Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,				
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,				
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,				
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,				
RA	Burtis K.C., Busam D.A., Butler H., Cadieu L.B., Davies P.,				
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,				
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,				
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,				
RA	Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,				
RA	Foster C., Gabrielian A.E., Gang N.S., Gelbart W.M., Glasser K.,				
RA	Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,				
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,				
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,				
RA	Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,				
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,				



RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Metkulov G., Mishina N.V., Kobayashi C., Morris J., Moshel A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shee B.C., Sidman K., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Stadler A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
DR EMBL; AF003698; AAF54888.2; -  
DR FlyBase; FBgn0038108; CG7518.  
DR InterPro; IPR001005; Myb\_DNA\_bind.  
DR PROSITE; PS00037; MYB\_1; UNKNOWN\_1.  
SQ SEQUENCE 2451 AA; 266959 MW; 088A2293F27481E2 CRC64;

Query Match 56.5%; Score 118; DB 5; Length 2451;  
Best Local Similarity 75.7%; Pred. No. 0.014;  
Matches 28; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 3 AAAAAAAAAAAAAAAAAAKKKKKKKKKKK 39  
DB 1374 AAAAAAAAAAAAAAAAAEAAAEKAKKKKKQAKK 1410

RESULT 2  
ID O35807 PRELIMINARY; PRT; 129 AA.  
AC O35807;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE MICROVASCULAR ENDOTHELIAL DIFFERENTIATION PROTEIN 2.  
GN MDG2.  
OS *Rattus norvegicus* (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=EPIDIDYMIS;  
RX MEDLINE=98172708; PubMed=9511718;  
RA Proels F., Loser B., Marx M.;  
RT "Differential expression of osteopontin, PC4, and CECS, a novel mRNA  
species, during in vitro angiogenesis.";  
RL Exp. Cell Res. 239:1-10(1998).  
DR EMBL; Y08769; CAA70022.1; -  
DR InterPro; IPR000719; Euk\_Pkinase.  
DR Pfam; PF00069; Pkinase; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR ATP-binding; Transferase.  
SQ SEQUENCE 129 AA; 15080 MW; 38102272BEEZEDB4 CRC64;

Query Match 50.2%; Score 105; DB 11; Length 129;  
Best Local Similarity 95.5%; Pred. No. 0.013;  
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 24 AAKKKKKKKKKKKKKKKKKKK 45  
DB 83 ASKKKKKKKKKKKKKKKKKK 104

RESULT 3  
O9H607  
PRELIMINARY; PRT; 517 AA.

ID O9H607 PRELIMINARY; PRT; 720 AA.  
AC O9H607;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE CDNA: FLJ21979 FIS, CLONE HEP06065 (FRAGMENT).  
OS *Homo sapiens* (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kawabata A., Hikiji T., Kodatate N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Ota T., Suzuki Y., Oshayashi M., Nishi T., Shibahara T.,  
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
RT "NEO human cDNA sequencing project.";  
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AK025632; BAB15196.1; -  
DR NON\_TER 720  
FT 720  
SQ SEQUENCE 720 AA; 84029 MW; A86586FEAA953D0B CRC64;

Query Match 50.2%; Score 105; DB 4; Length 720;  
Best Local Similarity 61.8%; Pred. No. 0.061;  
Matches 21; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

OY 12 AAAAAAAAAAAAAAAAAAKKKKKKKKKKKKK 45  
DB 678 AKKINSNDIVSISKKKKKKKKKKKKKKKK 711

RESULT 4  
ID O9L82 PRELIMINARY; PRT; 467 AA.  
AC O9L82;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE CBF5.  
OS *Euglena gracilis*.  
OC Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.  
OX NCBI\_TaxID=3039;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-STREPTOMYCIN-BLEACHED STRAIN;  
RX MEDLINE=20330353; PubMed=10871366;  
RA Watanabe Y., Gray M.W.;  
RT "Evolutionary appearance of genes encoding proteins associated with  
box H/ACA snoRNAs: Cbf5p in *Euglena gracilis*, an early diverging  
eukaryote, and candidate Garlp and Nop1op homologs in  
archaeobacteria.";  
RL Nucleic Acids Res. 28:2342-2352(2000).  
DR EMBL; AF234319; AAF77119.1; -  
DR InterPro; IPR002478; PUA.  
DR InterPro; IPR002501; TRUB\_N.  
DR Pfam; PF01472; PUA; 1.  
DR SMART; SM00359; TRUB\_N; 1.  
SQ SEQUENCE 467 AA; 52385 MW; 401089B66507A7B CRC64;

Query Match 48.8%; Score 102; DB 10; Length 467;  
Best Local Similarity 55.8%; Pred. No. 0.075;  
Matches 24; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAAKKKKKKKKKKKKK 44  
DB 420 EELAYAAAKRREREAAGDEKAKKKKKKKKK 462

RESULT 5  
O9LXR2 PRELIMINARY; PRT; 517 AA.

AC Q9LXR2;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
DE HYPOTHETICAL 59.7 KDA PROTEIN.  
GN T20N10-250.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,  
RA Rudd S., Lemcke K., Mayer K.F.X., Queller F., Salanoubat M.,  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL353032; CAB8307.1; -  
DR InterPro: IPR001810; F-box.  
DR Pfam: PF00646; F-box; 1.  
DR SMART: SM00256; FBOX; 1.  
DR PROSITE: PS50181; FBOX; 1.  
DR Hypothetical protein.  
SQ SEQUENCE 517 AA; 59689 MW; EC6D957D01F86E70 CRC64;  
Query Match 48.8%; Score 102; DB 10; Length 517;  
Best Local Similarity 80.8%; Pred. No. 0.082; Mismatches 0; Gaps 0;  
Matches 21; Conservative 0; Indels 0; Gaps 0;  
QY 20 EAAAAA  
DB 440 ERVGGKKKKKKKKKKKKKKKKKKKK 465  
RESULT 6  
ID 014347 PRELIMINARY; PRT; 354 AA.  
AC 014347;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE PUTATIVE ZUOTIN-LIKE PROTEIN C30D10.01 (FRAGMENT).  
GN SP3C30D10.01.  
OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-972;  
RA Wood V., Rajandream M.A., Barrell B.G., Lauber J., Hilbert H.,  
RA Duesterhoeft A.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: Z-DNA BINDING PROTEIN. COULD BE INVOLVED IN CHROMOSOME  
CC ORGANIZATION (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -1- SIMILARITY: CONTAINS A DNAJ-LIKE DOMAIN.  
DR HSSP: P25685; 1HDJ.  
DR InterPro: IPR001623; DnaJ\_N.  
DR Pfam: PF00226; DnaJ; 1.  
DR SMART: SM00271; DnaJ; 1.  
DR PROSITE: PS00636; DnaJ\_1; 1.  
DR PROSITE: PS50076; DnaJ\_2; 1.  
KW Hypothetical protein; Chaperone; DNA-binding; Nuclear protein.  
FT NON\_TER 1  
FT DOMAIN 10 81 DnaJ-LIKE.  
FT DOMAIN 217 267 ALA/LYS-RICH.

SQ SEQUENCE 354 AA; 40290 MW; 6071B58A3B60F558 CRC64;  
Query Match 48.6%; Score 101.5; DB 3; Length 354;  
Best Local Similarity 56.0%; Pred. No. 0.064;  
Matches 28; Conservative 3; Mismatches 12; Indels 7; Gaps 1;  
QY 2 EAAAAA-----EAAAAAEAAAAA  
DB 221 EAAAAAOKKKEEERAAAEAAAKASAAANKAKEDKKAKGRKKYVK 270  
RESULT 7  
ID 0951V6 PRELIMINARY; PRT; 531 AA.  
AC 0951V6;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DE HYPOTHETICAL 61.4 KDA PROTEIN (FRAGMENT).  
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecoidea; Macaca.  
OX NCBI\_TaxID=9541;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-TESTIS;  
RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,  
RA Tero K., Sugano S.;  
RT Isolation of novel full-length cDNA clones from macaque testis cDNA  
RT libraries.  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB071085; BAB64479.1; -  
KW Hypothetical protein.  
FT NON\_TER 531  
FT SEQUENCE 531 AA; 61389 MW; B5596B4F5CDD60C CRC64;  
Query Match 48.3%; Score 101; DB 6; Length 531;  
Best Local Similarity 95.2%; Pred. No. 0.1;  
Matches 20; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 25 AKKKKKKKKKKKKKKKKKKKKK 45  
DB 501 SKKKKKKKKKKKKKKKKKKKKK 521  
RESULT 8  
ID 09P529 PRELIMINARY; PRT; 128 AA.  
AC 09P529;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DE HYPOTHETICAL 15.2 KDA PROTEIN.  
GN B24H17.160.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Schulte U., Algn V., Hohnsels J., Brandt P., Fartmann B., Holland R.,  
RA Nyakatura G., Mewes H.W., Mannhauf G.;  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA German Neurospora genome project;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL356815; CAB92638.2; -  
KW Hypothetical protein.  
SQ SEQUENCE 128 AA; 15157 MW; 8C7C65C3DFB70765 CRC64;

Query Match 47.8%; Score 100; DB 3; Length 128;  
Best Local Similarity 100.0%; Pred. No. 0.035;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 KKKKKKKKKKKKKKKKK 45  
DB 71 KKKKKKKKKKKKKKKKK 90

RESULT 9  
AC Q9H5V6 PRELIMINARY; PRT; 168 AA.  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE CDNA: FLJ22976 FIS, CLONE KAT11222 (FRAGMENT).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,  
RA Suzuki Y., Ohtsuyoshi M., Nishi T., Shibahara T., Tanaka T.,  
RA Nakamura Y., Isegai T., Sugano S.;  
RT "NDO human cDNA sequencing project."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK026029; BAB15513.1; -  
FT NON\_TER 168  
SQ SEQUENCE 168 AA; 19549 MW; A19DBD195F8A1A90 CRC64;

Query Match 47.8%; Score 100; DB 4; Length 168;  
Best Local Similarity 100.0%; Pred. No. 0.044;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 KKKKKKKKKKKKKKKKK 45  
DB 140 KKKKKKKKKKKKKKKKK 159

RESULT 10  
AC Q64075 PRELIMINARY; PRT; 215 AA.  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE NUCLEOPORIN P62 HOMOLOG PROTEIN (FRAGMENT).  
OS Rattus sp.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Solurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10118;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE:95151924; Pubmed-7849178;  
RA Wang Z.Q., Akmal K.M., Kim K.H.;  
RT "An unusual nucleoporin-related messenger ribonucleic acid is present  
in the germ cells of rat testis";  
RL Biol. Reprod. 51:1022-1030(1994).  
DR EMBL; S75997; AAB3384.1; -  
FT NON\_TER 1  
SQ SEQUENCE 215 AA; 24593 MW; 098251C97A8FBD88 CRC64;

Query Match 47.8%; Score 100; DB 11; Length 215;  
Best Local Similarity 100.0%; Pred. No. 0.055;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 KKKKKKKKKKKKKKKKK 45  
DB 35 KKKKKKKKKKKKKKKKK 54

RESULT 11  
AC Q9LG29 PRELIMINARY; PRT; 260 AA.  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GENOMIC DNA, CHROMOSOME 3, BAC CLONE:FLD9.  
OS Arabidopsis thaliana (mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COL-0; BAC;  
RA Nakamura Y.;  
RT "Structural Analysis of Arabidopsis thaliana Chromosome 3. III."  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP002460; BAA97098.1; -  
DR InterPro: IPR001386; Linker\_histone.  
DR PRINTS: PRO0624; HISTONEH5.  
SQ SEQUENCE 260 AA; 33307 MW; 43E294C58131143 CRC64;

Query Match 47.8%; Score 100; DB 10; Length 260;  
Best Local Similarity 100.0%; Pred. No. 0.065;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 KKKKKKKKKKKKKKKKK 45  
DB 7 KKKKKKKKKKKKKKKKK 26

RESULT 12  
AC Q9NT34 PRELIMINARY; PRT; 380 AA.  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE HYPOTHEICAL 42.7 KDA PROTEIN (FRAGMENT).  
GN DKFP43411120.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=TESTIS;  
RA Ottenwelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL137556; CAB70810.1; -  
RT Hypothetical protein.  
FT NON\_TER 380  
SQ SEQUENCE 380 AA; 42689 MW; 67F5DD010346AFB CRC64;

Query Match 47.8%; Score 100; DB 4; Length 380;  
Best Local Similarity 100.0%; Pred. No. 0.092;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 KKKKKKKKKKKKKKKKK 45  
DB 355 KKKKKKKKKKKKKKKKK 374

RESULT 13





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:19:24 ; Search time 98.97 Seconds

(without alignments)  
50.503 Million cell updates/sec

Title: US-09-461-684-4

Perfect score: 209  
Sequence: 1 CEAATAAAATAAAATAAAE.....KKKKKKKKKKKKKKKKKK 45

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

A\_Geneseq\_032802.\*  
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	209	100.0	45	21	AA13783
2	119	56.9	36	21	AA13783
3	119	56.9	630	21	AA13783
4	119	56.9	640	21	AA13783
5	118	56.5	2451	22	AA13783
6	115.5	55.3	123	22	AA13783
7	115	55.0	123	22	AA13783
8	111	53.1	63	22	AA13783
9	110	52.6	141	22	AA13783
10	109	52.2	25	21	AA13781
11	109	52.2	59	21	AA13781

12	109	52.2	59	21	AA13783	Nuclear ligand for
13	109	52.2	59	22	AA13783	Nuclear ligand #2
14	109	52.2	59	22	AA13783	Nuclear acid trans
15	109	52.2	74	22	AA13783	Human polypeptide
16	109	52.2	75	22	AA13783	Human polypeptide
17	109	52.2	112	22	AA13783	Human polypeptide
18	109	52.2	150	22	AA13783	Human polypeptide
19	108	51.7	80	22	AA13783	Human polypeptide
20	108	51.7	126	22	AA13783	Human polypeptide
21	107	51.2	48	22	AA13783	Human polypeptide
22	107	51.2	63	22	AA13783	Human polypeptide
23	107	51.2	83	22	AA13783	Human polypeptide
24	106	50.7	26	22	AA13783	Human polypeptide
25	106	50.7	57	22	AA13783	Human polypeptide
26	106	50.7	64	22	AA13783	Human polypeptide
27	106	50.7	74	22	AA13783	Human polypeptide
28	106	50.7	74	22	AA13783	Human polypeptide
29	106	50.7	128	22	AA13783	Human polypeptide
30	106	50.7	272	22	AA13783	Human polypeptide
31	105	50.2	27	21	AA13783	Human polypeptide
32	105	50.2	31	22	AA13783	Human polypeptide
33	105	50.2	62	20	AA13783	Human polypeptide
34	105	50.2	62	22	AA13783	Human polypeptide
35	105	50.2	66	22	AA13783	Human polypeptide
36	105	50.2	66	22	AA13783	Human polypeptide
37	105	50.2	72	22	AA13783	Human polypeptide
38	105	50.2	83	22	AA13783	Human polypeptide
39	105	50.2	84	22	AA13783	Human polypeptide
40	105	50.2	122	22	AA13783	Human polypeptide
41	105	50.2	155	22	AA13783	Human polypeptide
42	105	50.2	175	11	AA13783	Human polypeptide
43	105	50.2	630	21	AA13783	Human polypeptide
44	105	50.2	640	21	AA13783	Human polypeptide
45	104	49.8	28	22	AA13783	Human polypeptide

## ALIGNMENTS

RESULT	1
AA13783	AA13783 standard; peptide: 45 AA.
XX	AA13783;
XX	10-NOV-2000 (first entry)
DE	Soluble tandem pEA/ PK peptide conjugate.
XX	PK peptide: cytosolic; vaccine: cytotoxic T cell; CTL; immunotherapy;
KM	major histocompatibility complex class I; MHC class I; antigen; tumour;
KM	prostate; breast; multiple myeloma; pEA peptide.
XX	Unidentified.
OS	Unidentified.
XX	WO200035949-A1.
PN	22-JUN-2000.
PD	14-DEC-1999; 99WO-US28724.
PF	14-DEC-1999; 98US-0113324.
PR	14-DEC-1998; 98US-0113324.
PA	(DEND-) DENDREON CORP.
XX	Laus R, Hakim I, Vidovic D;
XX	WPI; 2000-442365/38.
DR	Antigen modified by the covalent addition of a peptide that
XX	facilitates entry into antigen presenting cells, useful for producing
PT	compositions for immunizing against tumors and pathogens -
PT	



```

ID  AAB23593 standard; Protein; 640 AA.
XX
XX  AAB23593;
AC
XX
DT  11-JAN-2001 (first entry)
XX
XX  Modified fibre protein encoded in pW66.7R-F/ask21MSHB.
DE
XX  Viral vector; melanocyte-stimulating hormone receptor; MSH; cytostatic;
KM  tumour; malignant melanoma; fibre protein.
XX
OS  Synthetic.
XX  WO200050618-A1.
XX  31-AUG-2000.
XX  24-FEB-2000; 2000WO-JP01069.
XX  24-FEB-1999; 99JP-0093263.
XX  (NICA-) JAPANESE FOUND CANCER RES.
XX  Hamada H;
XX  WPI; 2000-549414/50.
XX  N-PSDB; AAA93825.
XX  Virus vector useful in diagnosis and treatment of tumors particularly
PT  malignant melanoma; constructed by fusing viral protein with ligand
PT  binding specifically to melanocyte-stimulating hormone receptor
XX
XX  Claim 24; Page 131-136; 145pp; Japanese.
XX
XX  This invention relates to a viral vector constructed by the fusion of a
CC  viral protein with a ligand which binding specifically to the
CC  melanocyte-stimulating hormone (MSH) receptor. The vector contains one
CC  of four linkers represented by sequences AAA93815-A93818 and
CC  AAB23583-B23586, and DNA encoding a fibre protein selected from those
CC  represented by AAA93819-A93826 and AAB23587-B23594. The vector has
CC  cytostatic activity, and can be used for gene therapy and in the
CC  diagnosis and treatment of tumours, particularly malignant melanomas. The
CC  present sequence represents a modified fibre protein used in the
CC  construction of the vector.
XX
XX  Sequence 640 AA:
SQ

```

Query Match 56.9%; Score 119; DB 21; Length 640;  
 Best Local Similarity 70.6%; Pred. No. 0.00016;  
 Matches 24; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

```

QY  12 AAEAAAAAEAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
DB  583 saasasaaagsakkkkkkkkkkkkkkkkkkkkkkkkkkkk 616

```

RESULT 5  
 ABB71574  
 ID ABB71574 standard; Protein; 2451 AA.

26-MAR-2002 (first entry)  
 Drosophila melanogaster polypeptide SEQ ID NO 41514.  
 Drosophila; developmental biology; cell signalling; insecticide;  
 pharmaceutical.  
 Drosophila melanogaster.  
 WO200171042-A2.

```

XX  27-SEP-2001.
XX
XX  23-MAR-2001; 2001WO-US09231.
XX
XX  23-MAR-2000; 2000US-191637P.
XX  11-JUL-2000; 2000US-0614150.
XX
XX  (PEKE ) PE CORP NY.
XX
XX  Venter JC, Adams M, Li PWD, Myers EW;
PI  WPI; 2001-656860/75.
XX  N-PSDB; ABL15677.
XX
XX  New isolated nucleic acid detection reagent for detecting 1000 or more
PT  genes from Drosophila and for elucidating cell signalling and cell-cell
PT  interactions -
XX
XX  Disclosure; SEQ ID NO 41514; 21pp + Sequence Listing; English.
XX
XX  The invention relates to an isolated nucleic acid detection reagent
CC  capable of detecting 1000 or more genes from Drosophila. The invention is
CC  useful in developmental biology and in elucidating cell signalling and
CC  cell-cell interactions in higher eukaryotes for the development of
CC  insecticides, therapeutics and pharmaceutical drugs. The invention
CC  discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
CC  sequences (AB101840-ABL16175) and the encoded proteins
CC  (AB57737-AB572072).
XX  The sequence data for this patent did not form part of the printed
CC  specification, but was obtained in electronic format directly from WIPO
CC  at ftp.wipo.int/pub/published_pct_sequences.
XX
XX  Sequence 2451 AA:
SQ

```

Query Match 56.5%; Score 118; DB 22; Length 2451;  
 Best Local Similarity 75.7%; Pred. No. 0.00072;  
 Matches 28; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

```

QY  3 AAAAAAEAAAAAEAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
DB  1374 aaaaaaataaaaaaataaaaaaataaaagkaklknkkgakx 1410

```

RESULT 6  
 AAO01368  
 ID AAO01368 standard; Protein; 123 AA.

AAO01368;  
 06-NOV-2001 (first entry)  
 Human polypeptide SEQ ID NO 15260.  
 Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 vaccine; peptide therapy; stem cell growth factor; hematopoiesis;  
 tissue growth factor; immunomodulatory; cancer; leukaemia;  
 nervous system disorders; arthritis; inflammation.  
 Homo sapiens.  
 WO200164835-A2.  
 07-SEP-2001.  
 26-FEB-2001; 2001WO-US04927.  
 28-FEB-2000; 2000US-0515126.  
 18-MAY-2000; 2000US-0577409.  
 (HYSE-) HYSEQ INC.





PR	26-JUL-2000	2000US-0220966
PR	16-AUG-2000	2000US-0220964
PR	14-AUG-2000	2000US-0224518
PR	14-AUG-2000	2000US-0224519
PR	14-AUG-2000	2000US-0225213
PR	14-AUG-2000	2000US-0225214
PR	14-AUG-2000	2000US-0225266
PR	14-AUG-2000	2000US-0225267
PR	14-AUG-2000	2000US-0225268
PR	14-AUG-2000	2000US-0225270
PR	14-AUG-2000	2000US-0225457
PR	14-AUG-2000	2000US-0225471
PR	14-AUG-2000	2000US-0225758
PR	14-AUG-2000	2000US-0225759
PR	18-AUG-2000	2000US-0225279
PR	22-AUG-2000	2000US-0226861
PR	22-AUG-2000	2000US-0226868
PR	23-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227109
PR	30-AUG-2000	2000US-0228292
PR	01-SEP-2000	2000US-0228287
PR	01-SEP-2000	2000US-0228343
PR	01-SEP-2000	2000US-0228344
PR	01-SEP-2000	2000US-0228344
PR	05-SEP-2000	2000US-0228519
PR	05-SEP-2000	2000US-0228513
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0233080
PR	08-SEP-2000	2000US-0233081
PR	12-SEP-2000	2000US-0233968
PR	14-SEP-2000	2000US-0233997
PR	14-SEP-2000	2000US-0233998
PR	14-SEP-2000	2000US-0233999
PR	14-SEP-2000	2000US-0234000
PR	14-SEP-2000	2000US-0234001
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234274
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	25-SEP-2000	2000US-0234998
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0235837
PR	29-SEP-2000	2000US-0235837
PR	29-SEP-2000	2000US-0235868
PR	29-SEP-2000	2000US-0235869
PR	02-OCT-2000	2000US-0235802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239937
PR	13-OCT-2000	2000US-0239937
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241121
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241826
PR	01-NOV-2000	2000US-0244617
PR	08-NOV-2000	2000US-0246674

CC	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246609
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	08-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0249207
PR	17-NOV-2000	2000US-0249208
PR	17-NOV-2000	2000US-0249210
PR	17-NOV-2000	2000US-0249211
PR	17-NOV-2000	2000US-0249212
PR	17-NOV-2000	2000US-0249213
PR	17-NOV-2000	2000US-0249214
PR	17-NOV-2000	2000US-0249215
PR	17-NOV-2000	2000US-0249216
PR	17-NOV-2000	2000US-0249217
PR	17-NOV-2000	2000US-0249244
PR	17-NOV-2000	2000US-0249245
PR	17-NOV-2000	2000US-0249264
PR	17-NOV-2000	2000US-0249265
PR	17-NOV-2000	2000US-0249297
PR	17-NOV-2000	2000US-0249299
PR	01-DEC-2000	2000US-0250160
PR	01-DEC-2000	2000US-0250391
PR	05-DEC-2000	2000US-0251030
PR	05-DEC-2000	2000US-0251088
PR	05-DEC-2000	2000US-0256719
PR	06-DEC-2000	2000US-0251479
PR	08-DEC-2000	2000US-0251856
PR	08-DEC-2000	2000US-0251869
PR	08-DEC-2000	2000US-0251889
PR	08-DEC-2000	2000US-0251990
PR	11-DEC-2000	2000US-0254097
PR	05-JAN-2001	2001US-0256978
XX		
PA	(HUMA-)	HUMAN GENOME SCI INC.
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX	WPI; 2001-465557/50.	
DR	N-PSDB; AAS29151.	
XX		
PT	Nucleic acid molecules encoding human secreted chromosomal binding	
PT	proteins, used in preventing, treating or ameliorating a disorder, e.g	
PT	Alzheimer's and Parkinson's diseases and cancers -	
XX		
PS	Claim 11; SEQ ID NO 260; 561pp; English.	
XX		
CC	The present invention relates to the isolation of novel DNA-binding	
CC	proteins, and cDNA (AAS29030-AAS29157) and genomic sequences encoding	
CC	for these proteins. DNA-binding proteins such as histones, chromo	
CC	(chromatin organisation modifier) domain proteins, and T-box binding	
CC	proteins may contribute to diseases resulting from aberrant DNA	
CC	organisation and/or gene transcription. The sequences of the invention	
CC	are useful in screening assays to identify antagonists and/or agonists	
CC	that may enhance or block activities mediated by DNA-binding proteins.	
CC	Blockers of DNA-binding proteins may be useful in treating disorders	
CC	such as malignant diseases (e.g. cancer), autoimmune disorders	
CC	(e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid	
CC	arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious	

CC diseases (e.g. HIV and neurological disorders (e.g. Alzheimer's  
CC disease)). The polynucleotide sequences of the invention may also be  
CC used in gene therapy. AM18154-AM18281 represent novel DNA-binding  
CC proteins.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pubdb/published\\_pcr\\_sequences](http://wipo.int/pubdb/published_pcr_sequences).  
XX  
XX  
Sequence 63 AA;  
SQ

Query Match	53.1%	Score 111;	DB 22;	Length 63;
Best Local Similarity	62.5%	Pred. No. 0.00011;		
Matches 25; Conservative	2;	Mismatches 13;	Indels 0;	Gaps 0

```
OY      6 AAAAAAAAAAAAAAKKKKKKKKKKKKKKKKKK 45
        | || : | | ||||| ||||| |||||
Db      9 adslaasqvaemvelkkkkkkkkkkkkkkkkk 48
```

RESULT	9
ABG26718	
ID	ABG26718 standard; Protein; 141 AA
xy	

DT 18-FEB-2002 (first entry)

Novel human diagnostic protein #26709

KW Human; chromosome mapping; gene mapping; gene therapy; forensic  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 YX

OS Homo sapiens.

PN WO200175067-A2

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631  
YY

PR	31-MAR-2000; 2000US-0540217
PR	03-APR-2000; 2000US-0640367

XX  
XX  
DIRECTOR, FBI

PI Drmanac RT, Liu C, Tang YT, xv

DR WPI; 2001-639362/73.

[illegible]

PT diagnostics, forensics, gene mapping, identification of mutations

PT biodiversity

PS Claim 20; SEQ ID No 57077; 103pp; English

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences, (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AB000010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPD  
CC at [http://wipd.int/pub/published\\_pct\\_sequences](http://wipd.int/pub/published_pct_sequences).  
XX  
XX Sequence 141 AA:  
XX

Query Match	52.6%	Score 110;	DB 22;	Length 141;
Best Local Similarity	53.3%	Pred. No. 0.0029;		
Matches 24; Conservative	2;	Mismatches 19;	Indels 0;	Gaps 0

```
QY      1 CEAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKKKKK 45
          | : | | | | | | | | | | | | | | | | | | |
Db      60 cttlvastrlksmveievarprkklklklklklklklklklkl 10
```

```

RESULT 10
AAB13781
ID AAB13781 standard; peptide; 25 AA
XX

```

AC AAB13781  
yy

DT 10-NOV-2000 (first entry,  
yy

DE	Soluble peptide antigen pEA
Y	

KW pEA peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy  
 KW major histocompatibility complex class 1; MHC class 1; antigen; tumour  
 KW prostate; breast; multiple myeloma.

Unidentified

PN WO200035949-A1

PD 22-JUN-2000.

PF 14-DEC-1999; 99WO-US29724

PR 14-DEC-1998; 98US-0112324

PA (DEND-) DENDREON CORP

PI Laus R, Hakim I, Vidovic D;

WPI; 2000-442365/38

PT Antigens modified by the covalent addition of a peptide that

PT compositions for immunizing against tumors and pathogens .

PS Claim 2; Page 26; 34pp; English

The present invention relates to compositions of modified soluble proteins or antigens capable of eliciting an enhanced *in vivo* cytotoxic T cell (CTL) response, i.e. a major histocompatibility complex (MHC) class I molecule response. The protein antigen is modified by the covalent addition of a peptide sequence which facilitates entry of the antigen into antigen presenting cells (APCs). The present sequence is one such peptide sequence which can be used to modify the soluble antigens. The present sequence is peptide PEA. The modified antigen composition may be used for immunising against, or treating a tumour e.g. prostate and breast carcinoma or multiple myeloma, or pathogen in mammals.

Query Match	52.2%	Score 109;	DB 21;	Length 25;
Best Local Similarity	100.0%	Pred. No. 7.1e-05;		
Matches 25; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

OY 1 CEAHAADPAHAHAHAHAHAHAHAHA 25  
 DB 1 CEAHAHAHAHAHAHAHAHAHAHAHA 25

## RESULT 11

AA98495  
 ID AAY98495 standard; Peptide: 59 AA.  
 AC AAY98495;  
 XX  
 DT 31-JUL-2000 (first entry)  
 DE  
 XX Nuclear ligand used in nucleic acid transporter system.  
 DE Transporter system; nucleic acid delivery; gene therapy; cancer;  
 KM carcinogenesis; cardiovascular disease; infection.  
 XX  
 OS Synthetic.  
 XX  
 PN US6033884-A.  
 PD 07-MAR-2000.  
 XX  
 PF 14-DEC-1993; 93US-0167641.  
 XX  
 PR 20-MAR-1992; 92US-0855389.  
 PR 19-MAR-1993; 93WO-US02725.  
 XX  
 PA (BAYU ) BAYLOR COLLEGE MEDICINE.  
 XX  
 PI Gottchalk S, Sparrow J, Cristiano RJ, Woo SLC, Smith LC;  
 DR WPI: 2000-281993/24.  
 XX  
 PT System for transporting nucleic acid into cells; useful e.g. in gene  
 PT therapy and for generating transgenic animals; comprises binding agent  
 PT linked to nucleic acid, surface ligand and lytic agent.  
 PS Claim 16; Column 123-124; 108pp; English.  
 XX  
 CC The present invention relates to a transporter system for delivering  
 CC nucleic acid to a cell. The system comprises a nucleic acid binding  
 CC complex, consisting of a binding molecule bonded non-covalently to the  
 CC nucleic acid, and covalently to a surface ligand, and a lytic agent. The  
 CC binding molecule is a spermine or a spermidine derivative. Nucleotide  
 CC sequences AA3633-A3652 and peptide sequences AAY98456-Y98500 are used  
 CC in the construction of the transporter system of the invention. The  
 CC transporter system is used in gene therapy, particularly to deliver  
 CC nucleic acids to hepatocytes, muscle cells or bone forming cells, e.g.  
 CC for treating cardiovascular disease, cancer, and infection. The  
 CC transporter systems are also used to create transgenic animals (as models  
 CC for human carcinogenesis or disease or for drug testing). Other uses  
 CC include transforming cells to produce proteins, or transfecting cells in  
 CC vitro to study the function of the nucleic acid. The use of a surface  
 CC ligand allows specific targeting of selected cells and tissues. The lytic  
 CC agent provides for release of the nucleic acid into the cellular  
 CC interior, from endosomes, without requiring endosomal or lysosomal  
 CC degradation.  
 XX  
 SQ Sequence 59 AA:

Query Match 52.2%; Score 109; DB 21; Length 59;  
 Best Local Similarity 88.5%; Pred. No. 0.00016;  
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 20 EAAAAAHHHHHHHHHHHHHHHHHHHH 45  
 DB 12 eapYkAKKKKKKKKKKKKKKKKKKK 37

## RESULT 12

AA59040  
 ID AAY59040 standard; peptide: 59 AA.  
 AC AAY59040;  
 XX  
 DT 07-MAR-2000 (first entry)  
 DE  
 XX Nuclear ligand for transporting nucleic acid to the nucleus.  
 DE  
 XX  
 DE Nucleic acid transport system; NTS; cell surface receptor; cytosol;  
 KM nuclear membrane; lysis moiety; transgenic animal; human disease;  
 KM nucleic acid delivery; cancer.  
 XX  
 OS Synthetic.  
 XX  
 PN US5994109-A.  
 PD 30-NOV-1999.  
 XX  
 PF 03-JUN-1995; 95US-0460890.  
 XX  
 PR 14-DEC-1993; 93US-0167641.  
 PR 20-MAR-1992; 92US-0855389.  
 PR 19-MAR-1993; 93WO-US02725.  
 PR 14-DEC-1993; 93US-0167641.  
 XX  
 PA (BAYU ) BAYLOR COLLEGE MEDICINE.  
 XX  
 PI Woo SLC, Cristiano RJ, Gottchalk S, Sparrow J, Smith LC;  
 DR WPI: 2000-038262/03.  
 XX  
 PT Nucleic acid transport system, useful for creating transgenic animals  
 PT for assessing human disease such as cancer in an animal model.  
 PS Disclosure; Columns 119-122; 107pp; English.

XX  
 CC The invention relates to a nucleic acid transport system (NTS) for  
 CC delivering nucleic acid into a cell. The NTS contains but is not limited  
 CC to 5 components: (a) the nucleic acid or a macromolecule to be delivered;  
 CC (b) a moiety that recognizes and binds to a cell surface receptor or  
 CC antigen or is capable of entering a cell through cytosol; (c) a nucleic  
 CC acid or macromolecular molecule binding moiety; (d) a moiety that is  
 CC capable of moving or initiating movement through a nuclear membrane; and/  
 CC (e) a lysis moiety that enables the transport of the entire complex  
 CC from the cell surface directly into the cytoplasm of the cell. The NTS  
 CC delivers nucleic acid into the cellular interior as well as the nucleus  
 CC of specific cells. The NTS can be used to treat disorders by targeting  
 CC specific nucleic acid accordingly. The NTS can also be used to create  
 CC transgenic animals for assessing human disease, such as cancer, in an  
 CC animal model. The NTS can be used in vitro with tissue culture cells  
 CC which allows the role of various nucleic acids to be studied by targeting  
 CC specific expression into specifically targeted tissue culture cells. The  
 CC lysis agent within the NTS avoids the problem of endosomal/lysosomal  
 CC degradation.  
 XX  
 SQ Sequence 59 AA:

Query Match 52.2%; Score 109; DB 21; Length 59;  
 Best Local Similarity 88.5%; Pred. No. 0.00016;  
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 20 EAAAAAHHHHHHHHHHHHHHHHHHHH 45  
 DB 12 eapYkAKKKKKKKKKKKKKKKKKKK 37

## RESULT 13

AA04285  
 ID AA04285 standard; Peptide: 59 AA.  
 AC AA04285;

XX 23-OCT-2001 (first entry)  
DT  
XX  
XX Nuclear ligand #2 used in nucleic acid transporter system.  
DE  
XX  
XX Nucleic acid transporter; cytosols; ligand; lysis agent; spacer molecule;  
KW gene therapy; hepatocyte; muscle; bone forming cell.  
XX  
XX  
OS Synthetic.  
XX  
XX US6177554-B1.  
PN  
XX  
XX 23-JAN-2001.  
PD  
XX  
XX 05-JUN-1995; 95US-0462040.  
PF  
XX  
XX 14-DEC-1993; 93US-0167641.  
PR 20-MAR-1992; 92US-085389.  
PR 19-MAR-1993; 93WO-US02725.  
XX  
XX  
XX (BAYU ) BAYLOR COLLEGE MEDICINE.  
PA  
XX  
XX Woo SLIC, Smith LC, Cristiano RJ, Gottchalk S, Sparrow J;  
PI  
XX  
XX WPI: 2001-365933/38.  
DR  
XX  
XX Nucleic acid transport system, useful for creating transgenic animals  
PT for assessing human disease such as cancer in an animal model .  
XX  
XX  
XX  
PS Disclosure: Column 8; 111pp; English.  
XX  
XX The sequence represents the nuclear ligand, #2, used in a nucleic acid  
CC transporter system. The nucleic acid transporter system uses nucleic acid  
CC binding complexes containing surface ligands which are capable of binding  
CC to a cell surface receptor and entering the cell through cytosols. The  
CC compounds of the invention are either ligands, binding molecules (surface  
CC ligands), lysis agents, spacer molecules or their intermediates. The  
CC nucleic acid transporter systems to deliver nucleic acid into specific  
CC cells e.g. in gene therapy to deliver nucleic acid into hepatocytes,  
CC muscle cells or bone forming cells.  
CC  
XX  
XX Sequence 59 AA:  
SQ

Query Match 52.2%; Score 109; DB 22; Length 59;  
Best Local Similarity 88.5%; Pred. No. 0.00016;  
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 20 EAAAKKKKKKKKKKKKKKKKKKKKK 45  
DB 12 eapYakKKKKKKKKKKKKKKKKKK 37

RESULT 14  
AAB45848  
ID AAB45848 standard; Protein: 59 AA.  
XX  
XX AAB45848;  
AC  
XX  
XX 21-MAR-2001 (first entry)  
DT  
XX  
XX Nucleic acid transporter system peptide ligand SEQ ID NO 60.  
DE  
XX  
XX Nucleic acid delivery; nucleic acid transporter system; hormone; enzyme;  
KW growth factor; clotting factor; apolipoprotein; receptor; drug; oncogene;  
KW tumor antigen; tumor suppressor; viral antigen; parasitic antigen;  
KW bacterial antigen.  
XX  
XX  
XX (unidentified.  
OS  
XX  
XX US6150168-A.  
PN  
XX

PD 21-NOV-2000.  
XX  
XX  
XX 05-JUN-1995; 95US-0460971.  
PF  
XX  
XX 14-DEC-1993; 93US-0167641.  
PR 20-MAR-1992; 92US-085389.  
PR 19-MAR-1993; 93WO-US02725.  
XX  
XX  
XX (BAYU ) BAYLOR COLLEGE MEDICINE.  
PA  
XX  
XX Gottchalk S, Sparrow J, Cristiano RJ, Smith LC, Woo SLIC;  
PI  
XX  
XX WPI: 2001-049093/06.  
DR  
XX  
XX Nucleic acid transporter system for delivering nucleic acid into a  
PT cell, useful for delivering proteins and polypeptides to cells,  
PT including growth factors, enzymes, hormones, and tumor suppressors  
XX  
XX  
XX Disclosure: Column 123-124; 105pp; English.  
XX  
XX This invention describes a novel system (I) for delivering a nucleic acid  
CC to a cell, comprising a binding complex comprising a ligand binding  
CC molecule noncovalently bound to a nucleic acid and covalently linked to a  
CC surface ligand, and a second binding complex comprising a second binding  
CC molecule noncovalently bound to a nucleic acid and covalently linked to a  
CC nuclear ligand. The complexes are simultaneously bound to the nucleic  
CC acid. The nucleic acid transporter system can also be used in a method  
CC for the in vivo targeting of the insertion of DNA into a cell. It can  
CC also be used in processes for producing transformed cell lines. The  
CC system can be used to deliver a variety of proteins and polypeptides,  
CC such as hormones, growth factors, enzymes, clotting factors,  
CC apolipoproteins, receptors, drugs, oncogenes, tumor antigens, tumor  
CC suppressors, viral antigens, parasitic antigens, and bacterial antigens.  
CC The transporter system uses lysis agents to overcome the problems of  
CC endosomal/lysosomal degradation seen with prior art systems.  
XX  
XX  
XX Sequence 59 AA:  
SQ

Query Match 52.2%; Score 109; DB 22; Length 59;  
Best Local Similarity 88.5%; Pred. No. 0.00016;  
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 20 EAAAKKKKKKKKKKKKKKKKKKKKK 45  
DB 12 eapYakKKKKKKKKKKKKKKKKKK 37

RESULT 15  
AAO03278  
ID AAO03278 standard; Protein: 74 AA.  
XX  
XX AAO03278;  
AC  
XX  
XX 06-NOV-2001 (first entry)  
DT  
XX  
XX Human polypeptide SEQ ID NO 17170.  
DE  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200164835-A2.  
PN  
XX  
XX 07-SEP-2001.  
PD  
XX  
XX 26-FEB-2001; 2001WO-US04927.  
PF  
XX  
XX 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.  
PR



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: July 1, 2002, 06:28:12 ; Search time 49.45 Seconds

(without alignments)  
85,499 Million cell updates/sec

Title: US-09-461-684-5

Perfect score: 243

Sequence: 1 GGFAGTAGFIENGWEGMID.....KKKKKKKKKKKKKKKKKK 44

## Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

1: PIR\_71:\*\*  
2: PIR:\*\*  
3: PIR:\*\*  
4: PIR:\*\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match length	ID	Description
1	134	55.1	550	1 HMTVS2	hemagglutinin prec
2	134	55.1	550	1 HMTVS3	hemagglutinin prec
3	134	55.1	550	1 HMTV77	hemagglutinin prec
4	134	55.1	550	1 HMTV80	hemagglutinin prec
5	134	55.1	550	1 HMTV33	hemagglutinin prec
6	134	55.1	550	1 HMTV89	hemagglutinin prec
7	134	55.1	550	1 HMTV21	hemagglutinin prec
8	134	55.1	550	1 HMTV98	hemagglutinin prec
9	134	55.1	550	1 HMTV15	hemagglutinin prec
10	134	55.1	550	2 JQ1153	hemagglutinin prec
11	134	55.1	550	2 JQ1154	hemagglutinin prec
12	134	55.1	550	2 JQ1155	hemagglutinin prec
13	134	55.1	550	1 HMTV4	hemagglutinin prec
14	134	55.1	550	1 HMTV4A	hemagglutinin prec
15	134	55.1	550	1 HMTV4M	hemagglutinin prec
16	134	55.1	550	1 HMTV4U	hemagglutinin prec
17	134	55.1	570	1 A45591	hemagglutinin prec
18	134	55.1	570	2 S22013	hemagglutinin prec
19	133	54.7	561	1 HMTV49	hemagglutinin prec
20	133	54.7	561	1 HMTV84	hemagglutinin prec
21	132	54.3	565	1 HMTVE1	hemagglutinin prec
22	132	54.3	565	1 HMTVE3	hemagglutinin prec
23	132	54.3	566	1 HMTVE	hemagglutinin prec
24	132	54.3	567	1 HMTV	hemagglutinin prec
25	131	53.9	562	2 S38637	hemagglutinin - in
26	131	53.9	550	1 HMTV86	hemagglutinin prec
27	131	53.9	560	1 HMTV77	hemagglutinin prec
28	131	53.9	565	1 HMTVE2	hemagglutinin prec
29	131	53.9	565	1 HMTVE4	hemagglutinin prec

30	131	53.9	565	1 HMTVE5	hemagglutinin prec
31	131	53.9	565	1 HMTVE6	hemagglutinin prec
32	131	53.9	565	1 HMTVE7	hemagglutinin prec
33	131	53.9	565	1 HMTVE8	hemagglutinin prec
34	131	53.9	565	1 HMTVE9	hemagglutinin prec
35	131	53.9	565	1 HMTVE	hemagglutinin prec
36	131	53.9	565	1 HMTVEE	hemagglutinin prec
37	131	53.9	565	1 S33703	hemagglutinin prec
38	131	53.9	570	2 S22014	hemagglutinin prec
39	131	53.9	570	2 S22015	hemagglutinin prec
40	131	53.9	570	2 S22016	hemagglutinin prec
41	131	53.9	570	2 S22017	hemagglutinin prec
42	131	53.9	570	2 S22018	hemagglutinin prec
43	131	53.9	570	2 S22019	hemagglutinin prec
44	131	53.9	570	2 S22020	hemagglutinin prec
45	131	53.9	570	2 S22021	hemagglutinin prec

## ALIGNMENTS

RESULT 1  
HMTVS2 hemagglutinin precursor - influenza A virus (strain A/swine/126/82) (fragment)  
C:Species: influenza A virus  
C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 18-Sep-1998  
C:Accession: A29971  
R:Kidd, H.; Shortridge, K.F.; Webster, R.G.  
Virology 162, 160-166, 1988  
A:Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China  
A:Reference number: A94370; MUID:88101364  
A:Accession: A29971  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M19056; NID:g324208  
A:Note: the sequence in Genbank entry FLAHP, release 106, (PID:g324209) differs fro  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HAI #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domains: transmembrane #status predicted <TM1>  
F:8-22-38-165-285-483/Binding site: carbohydrate (asn) (covalent) #status predicted  
F:114-166-52-277-64-76-139-473-201-305/Disulfide bonds: #status predicted  
F:535/546-349/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match  
Best Local Similarity 100.0%; Score 134; DB 1; Length 550;  
Pred. No. 3.1e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGFAGTAGFIENGWEGMIDGWTG 24  
DB 330 GGFAGTAGFIENGWEGMIDGWTG 352

RESULT 2  
HMTVS3 hemagglutinin precursor - influenza A virus (strain A/swine/81/78) (fragment)  
C:Species: influenza A virus  
C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 18-Sep-1998  
C:Accession: B29971  
R:Kidd, H.; Shortridge, K.F.; Webster, R.G.  
Virology 162, 160-166, 1988  
A:Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China  
A:Reference number: A94370; MUID:88101364  
A:Accession: B29971  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M19057; NID:g324210  
A:Note: the sequence in Genbank entry FLAHP, release 106, (PID:g324211) differs fro  
C:Genetics:

A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:300-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM>  
F:8-22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGMYG 24  
DB 330 GLFGAIGFIENGWEGMIDGMYG 352

RESULT 3  
HMTV77  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/5/77) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999  
C:Accession: A27813  
R:Kida, H.; Kawoka, Y.; Naewe, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: A27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16737; NID:g324081; PIDN:AAA43145.1; PID:g324082  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM>  
F:8-22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGMYG 24  
DB 330 GLFGAIGFIENGWEGMIDGMYG 352

RESULT 4  
HMTV80  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/8/80) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 18-Sep-1998  
C:Accession: B27813  
R:Kida, H.; Kawoka, Y.; Naewe, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: B27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16738; NID:g324083  
A:Note: The translation in Fig. 2 is inconsistent with the nucleotide sequence in Fig. 2  
C:Genetics:

A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM>  
F:8-22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGMYG 24  
DB 330 GLFGAIGFIENGWEGMIDGMYG 352

RESULT 5  
HMTV73  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/33/80) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999  
C:Accession: C27813  
R:Kida, H.; Kawoka, Y.; Naewe, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: C27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16739; NID:g324085; PIDN:AAA43145.1; PID:g324086  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM>  
F:8-22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGMYG 24  
DB 330 GLFGAIGFIENGWEGMIDGMYG 352

RESULT 6  
HMTV80  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/7/82) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999  
C:Accession: D27813  
R:Kida, H.; Kawoka, Y.; Naewe, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: D27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16740; NID:g324087; PIDN:AAA43146.1; PID:g324088  
C:Genetics:  
A:Map position: segment 4



C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1; status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2; status predicted <HA2>  
F:520-536/Domains: transmembrane; status predicted <TM>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3,1e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWTG 24  
|||||

Db 330 GLFGAIGFIENGWEGMIDGWTG 352

RESULT 7  
HMIIV21  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/21/82) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 18-Sep-1998  
C:Accession: E27813  
R:Kida, H.; Kawoka, Y.; Naeve, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: E27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16741; NID:g324089  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer  
F:1-328/Product: hemagglutinin HA1; status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2; status predicted <HA2>  
F:520-536/Domains: transmembrane; status predicted <TM>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3,1e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWTG 24  
|||||

Db 330 GLFGAIGFIENGWEGMIDGWTG 352

RESULT 8  
HMIIV98  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/9/85) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 18-Sep-1998  
C:Accession: F27813  
R:Kida, H.; Kawoka, Y.; Naeve, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: F27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16742; NID:g324091  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-328/Product: hemagglutinin HA1; status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2; status predicted <HA2>  
F:520-536/Domains: transmembrane; status predicted <TM>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3,1e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWTG 24  
|||||

Db 330 GLFGAIGFIENGWEGMIDGWTG 352

RESULT 9  
HMIIV15  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/10/85) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999  
C:Accession: G27813  
R:Kida, H.; Kawoka, Y.; Naeve, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: G27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16743; NID:g324093; PIDN:AAA3149.1; PID:g324094  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1; status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2; status predicted <HA2>  
F:520-536/Domains: transmembrane; status predicted <TM>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3,1e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWTG 24  
|||||

Db 330 GLFGAIGFIENGWEGMIDGWTG 352

RESULT 10  
JQ1153  
hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/7/75) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C>Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 20-Jun-2000  
C:Accession: JQ1153  
R:Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.  
J. Gen. Virol. 72, 2007-2010, 1991  
A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian  
A:Reference number: JQ1153; MUID:91341491  
A:Accession: JQ1153  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <YAS>  
A:Cross-references: GB:D00929; NID:g221279; PIDN:BA00769.1; PID:g221280  
A:Note: the authors translated the codon GGG for residue 218 as Glu  
A:Note: residues 528-533 are not shown in this publication  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; homotrimer  
F:1-328/Product: hemagglutinin HA1; status predicted <HA1>

F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 2; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3,1e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAAGFIENGWEGMIDGWYG 24  
Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

## RESULT 11

hemagglutinin precursor - influenza A virus (strain A/goose/Hong Kong/10/76) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: Influenza A virus  
C>Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 20-Jun-2000  
C:Accession: J01154  
R:Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.  
J. Gen. Virol. 72, 2007-2010, 1991  
A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3  
A:Reference number: J01153; MUID:91341491  
A:Accession: J01154  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <YAS>  
A:Cross-references: GB:D00930; NID:q221273; PIDN:AAA00771.1; PID:q221274  
A:Note: the authors translated the codon GCG for residue 218 as Glu  
C:Superfamily: Influenza virus hemagglutinin  
C:Keywords: glycoprotein; homotrimer  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 2; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3,1e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAAGFIENGWEGMIDGWYG 24  
Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

## RESULT 12

hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/64/76) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: Influenza A virus  
C>Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 20-Jun-2000  
C:Accession: J01155  
R:Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.  
J. Gen. Virol. 72, 2007-2010, 1991  
A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3  
A:Reference number: J01153; MUID:91341491  
A:Accession: J01155  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <YAS>  
A:Cross-references: GB:D00931; NID:q221277; PIDN:AAA00771.1; PID:q221278  
A:Note: the authors translated the codon GCG for residue 218 as Glu, GCC for residue 538  
C:Superfamily: Influenza virus hemagglutinin  
C:Keywords: glycoprotein; homotrimer  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>  
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 2; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3,1e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAAGFIENGWEGMIDGWYG 24  
Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

## RESULT 13

hemagglutinin precursor - influenza A virus  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: Influenza A virus  
C>Date: 28-Feb-1981 #sequence\_revision 28-Feb-1981 #text\_change 22-Oct-1999  
C:Accession: A93705; A93233; A04051; A93231; A94441  
R:Both, G.W.; Sleigh, M.J.  
Nucleic Acids Res. 8, 2561-2575, 1980  
A:Title: Complete nucleotide sequence of the haemagglutinin gene from a human influenza  
A:Reference number: A93705; MUID:81053698  
A:Accession: A93705  
A:Molecule type: genomic RNA  
A:Residues: 1-566 <BOT>  
A:Cross-references: GB:Y01103  
A:Experimental source: strain A/NT/60/68/29C  
A:Note: human influenza strain A/NT/60/68/29C is a laboratory-isolated variant of A/  
R/Doppeide, T.A.; Ward, C.W.  
FBS Lett. 110, 181-183, 1980  
A:Title: The disulphide bonds of a Hong Kong influenza virus hemagglutinin.  
A:Reference number: A91276; MUID:80179105  
A:Contents: annotation; disulfide bonds  
R:Gerling, M.J.; Bye, J.; Skehel, J.; Waterfield, M.  
Nature 287, 301-306, 1980  
A:Title: Cloning and DNA sequence of double-stranded copies of haemagglutinin genes  
A:Reference number: A93233; MUID:81030852  
A:Accession: A93233  
A:Molecule type: genomic RNA  
A:Residues: 1-24, 'S', '26', 'D', '28-159', 'G', '161-197', 'I', '199-241', 'L', '243-249' <GET>  
A:Experimental source: strain X-31[H3]  
C:Superfamily: Influenza virus hemagglutinin  
C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>  
F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>  
F:536-552/Domain: transmembrane #status predicted <TM1>  
F:30-482,68-293,80-92,155-489,297-321/Disulfide bonds: #status experimental  
F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 3,2e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAAGFIENGWEGMIDGWYG 24  
Db 346 GLFGAIAAGFIENGWEGMIDGWYG 368

## RESULT 14

hemagglutinin precursor - influenza A virus (strain A/Aichi/2/68)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: Influenza A virus  
C>Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 16-Jul-1999  
C:Accession: A93231; A04051  
R:Verhoeven, M.; Fang, R.; Min Jou, W.; Devos, R.; Huybrecock, D.; Saman, E.; Fiers  
Nature 286, 771-776, 1980  
A:Title: Antigenic drift between the haemagglutinin of the Hong Kong influenza strain  
A:Reference number: A93231; MUID:80254693  
A:Accession: A93231  
A:Molecule type: genomic RNA  
A:Residues: 1-566 <VER>  
A:Cross-references: GB:J02090; NID:q324131; PIDN:AAA43178.1; PID:q324132  
C:Superfamily: Influenza virus hemagglutinin  
C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>

F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>  
F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 3.2e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGMYG 24  
|||||  
Db 346 GLFGAIGFIENGWEGMIDGMYG 368

# RESULT 15

HMIYHM  
hemagglutinin precursor - influenza A virus (strain A/Mem/102/72)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C>Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 31-Mar-2000  
C:Accession: A94441; A04051  
R:Steigh, M.J.; Both, G.W.; Brownlee, G.G.; Bender, V.J.; Moss, B.A.  
in Structure and Variation in Influenza Virus, Laver, G., and Air, G., eds., pp.69-79, F  
A:Title: The haemagglutinin gene of influenza A virus: nucleotide sequence analysis of C  
A:Reference number: A94441  
A:Accession: A94441  
A:Molecule type: genomic RNA  
A:Residues: 1-566 <SDE>  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>  
F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>  
F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 3.2e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGMYG 24  
|||||  
Db 346 GLFGAIGFIENGWEGMIDGMYG 368

Search completed: July 1, 2002, 06:28:12  
Job time: 714 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:31:35 ; Search time 25.37 Seconds

(without alignments)  
67.153 Million cell updates/sec

Title: US-09-461-684-5

Perfect score: 243  
Sequence: 1 CGTGAAGTACGTGEGMDID.....KKKKKKKKKKKKKKKKKK 44

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 segs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	134	55.1	550	HEMA_IADH1	P12582 influenza a
2	134	55.1	550	HEMA_IADH2	P12583 influenza a
3	134	55.1	550	HEMA_IADH3	P12584 influenza a
4	134	55.1	550	HEMA_IADH4	P12585 influenza a
5	134	55.1	550	HEMA_IADH5	P12586 influenza a
6	134	55.1	550	HEMA_IADH6	P12587 influenza a
7	134	55.1	550	HEMA_IADH7	P12588 influenza a
8	134	55.1	550	HEMA_IADH8	P12589 influenza a
9	134	55.1	550	HEMA_IADH9	P12590 influenza a
10	134	55.1	550	HEMA_IADH10	P12591 influenza a
11	134	55.1	550	HEMA_IADH11	P12592 influenza a
12	134	55.1	550	HEMA_IADH12	P12593 influenza a
13	134	55.1	550	HEMA_IADH13	P12594 influenza a
14	134	55.1	550	HEMA_IADH14	P12595 influenza a
15	134	55.1	550	HEMA_IADH15	P12596 influenza a
16	134	55.1	550	HEMA_IADH16	P12597 influenza a
17	134	55.1	550	HEMA_IADH17	P12598 influenza a
18	134	55.1	550	HEMA_IADH18	P12599 influenza a
19	134	55.1	550	HEMA_IADH19	P12600 influenza a
20	134	55.1	550	HEMA_IADH20	P12601 influenza a
21	134	55.1	550	HEMA_IADH21	P12602 influenza a
22	134	55.1	550	HEMA_IADH22	P12603 influenza a
23	134	55.1	550	HEMA_IADH23	P12604 influenza a
24	134	55.1	550	HEMA_IADH24	P12605 influenza a
25	134	55.1	550	HEMA_IADH25	P12606 influenza a
26	134	55.1	550	HEMA_IADH26	P12607 influenza a
27	134	55.1	550	HEMA_IADH27	P12608 influenza a
28	134	55.1	550	HEMA_IADH28	P12609 influenza a
29	134	55.1	550	HEMA_IADH29	P12610 influenza a
30	134	55.1	550	HEMA_IADH30	P12611 influenza a
31	134	55.1	550	HEMA_IADH31	P12612 influenza a
32	134	55.1	550	HEMA_IADH32	P12613 influenza a
33	134	55.1	550	HEMA_IADH33	P12614 influenza a

34	131	53.9	565	HEMA_IADH34	P16996 influenza a
35	131	53.9	565	HEMA_IADH35	P16997 influenza a
36	131	53.9	565	HEMA_IADH36	P16998 influenza a
37	131	53.9	565	HEMA_IADH37	P16999 influenza a
38	131	53.9	565	HEMA_IADH38	P17000 influenza a
39	131	53.9	565	HEMA_IADH39	P17001 influenza a
40	131	53.9	565	HEMA_IADH40	P17002 influenza a
41	131	53.9	565	HEMA_IADH41	P17003 influenza a
42	131	53.9	565	HEMA_IADH42	P17004 influenza a
43	131	53.9	565	HEMA_IADH43	P17005 influenza a
44	131	53.9	565	HEMA_IADH44	P17006 influenza a
45	131	53.9	565	HEMA_IADH45	P17007 influenza a

## ALIGNMENTS

RESULT 1  
HEMA\_IADH1 STANDARD: PRT; 550 AA.  
AC P12582; Q84021; Q84022;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor (Contains: Hemagglutinin HAI chain;  
DE Hemagglutinin HAZ chain) (Fragment).  
GN HA.  
OS Influenza A virus (strain A/Duck/Hokkaido/5/77).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses;  
OC Influenza A virus.  
OX NCBI\_TaxID=11357;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=87265458; PubMed=2440178;  
RX Kida H., Kawakura Y., Naeve C.W., Webster R.G.;  
RT "Antigenic and genetic conservation of H3 influenza virus in wild  
RT ducks";  
RT virology 159:109-119(1987).  
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
CC or send an email to [license@sdb-sdb.ch](mailto:license@sdb-sdb.ch)).  
CC  
CC EMBL: M16737; AAA43143.1; -  
CC PIR: A27813; HMI1V77.  
CC InterPro: IPR001364; Hemagglutn.  
CC Pfam: PF00509; Hemagglutinin.1.  
CC ProDom: PD000225; Hemagglutn.1.  
KW Envelope protein; Hemagglutinin; glycoprotein.  
KW NON\_TER  
FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.  
FT CHAIN 330 550 HEMAGGLUTININ HAZ CHAIN.  
FT CARBOHYD 8 22 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 22 38 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 38 165 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 165 285 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 285 483 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 483 550 N-LINKED (GLCNAC... ) (POTENTIAL).  
SQ SEQUENCE 550 AA; 61705 MW; 7E7ACEF71FC969A CR64;

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 2.3e-07;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFTENGMEGMDGMYG 24  
Db 330 GLFGAIGFTENGMEGMDGMYG 352

## RESULT 2

HEMA\_IADH2 STANDARD; PRT; 550 AA.  
ID HEMA\_IADH2 STANDARD; PRT; 550 AA.  
AC P12583; Q84011; (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;  
DE Hemagglutinin HA2 chain] (Fragment).  
GN HA.  
OS Influenza A virus (strain A/Duck/Hokkaido/8/80).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses;  
OC Influenza A virus.  
OX NCBI\_TaxID=11358;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=67265458; PubMed=2440178;  
RA Kida H., Kawacka Y., Naeve C.W., Webster R.G.;  
RT "Antigenic and genetic conservation of H3 influenza virus in wild  
RT ducks";  
RL Virology 159:109-119(1987).  
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: M16738; AAA43144.1; -  
DR PIR: B27813; HMTV33  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR ProDom: PD000225; Hemagglutn; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON\_TER 1 1  
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
FT CARBOHYD 330 550 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 137 137 K -> N (IN PIR DATA BANK).  
SQ SEQUENCE 550 AA; 61659 MW; A107023ACC9CC353 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;

Best Local Similarity 100.0%; Pred. No. 2.3e-07;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## RESULT 3

HEMA\_IADH3

ID HEMA\_IADH3 STANDARD; PRT; 550 AA.

AC P12584; Q84012; Q89793;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;  
DE Hemagglutinin HA2 chain] (Fragment).

GN HA.

OS Influenza A virus (strain A/Duck/Hokkaido/33/80).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza virus A and B group; Influenza A viruses;

OC Influenza A virus.

OX NCBI\_TaxID=11359;  
RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=67265458; PubMed=2440178;

RA Kida H., Kawacka Y., Naeve C.W., Webster R.G.;

RT "Antigenic and genetic conservation of H3 influenza virus in wild  
RT ducks";

RL Virology 159:109-119(1987).

CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----

DR EMBL: M16739; AAA43145.1; -

DR PIR: C27813; HMTV33

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR ProDom: PD000225; Hemagglutn; 1.

KW Envelope protein; Hemagglutinin; Glycoprotein.

FT NON\_TER 1 1

FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.

FT CARBOHYD 330 550 HEMAGGLUTININ HA2 CHAIN.

FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 137 137 K -> N (IN PIR DATA BANK).

SQ SEQUENCE 550 AA; 61577 MW; 6C30BF67CFDCB7DE CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;

Best Local Similarity 100.0%; Pred. No. 2.3e-07;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## RESULT 4

HEMA\_IADH4

HEMA\_IADH4

STANDARD; PRT; 550 AA.

P12585; Q84013; Q84014;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;  
DE Hemagglutinin HA2 chain] (Fragment).

GN HA.

OS Influenza A virus (strain A/Duck/Hokkaido/7/82).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

```
CC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawacka Y., Naeye C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
  ducks.";
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
  CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
  (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M16740; AAA43146.1; -.
DR PIR: D27813; HMIY89.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61664 MW; A16B2CF8CBBD9D0 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWYG 24
DB 330 GLFGAIGFIENGWEGMIDGWYG 352

RESULT 5
HEMA_IADH5 STANDARD; PRT; 550 AA.
ID HEMA_IADH5 Q84015; Q84016;
AC P12586; Q84015; Q84016;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
  Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/21/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11361;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawacka Y., Naeye C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
  ducks.";
RL Virology 159:109-119(1987).
```

```
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
  CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
  (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M16741; AAA43147.1; -.
DR PIR: E27813; HMIY21.
DR HSSP: P03437; 5SMG.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 7 7 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 179 YV -> VI (IN PIR DATA BANK).
FT CONFLICT 388 388 K -> T (IN PIR DATA BANK).
SQ SEQUENCE 550 AA; 61856 MW; 48401C867A15B8C CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWYG 24
DB 330 GLFGAIGFIENGWEGMIDGWYG 352

RESULT 6
HEMA_IADH6 STANDARD; PRT; 550 AA.
ID HEMA_IADH6 Q84017;
AC P12587; Q84017;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
  Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/9/85).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11362;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawacka Y., Naeye C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
  ducks.";
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
  CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
  (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M16742; AAA3148.1; -  
 DR PIR: F27813; HMIY98.  
 DR InterPro: IPR001364; Hemagglutn.  
 DR Pfam: PF00509; Hemagglutinin; 1.  
 DR ProDom: PD000225; Hemagglutn; 1.  
 KW Envelope protein; Hemagglutinin; Glycoprotein.  
 FT NON\_TER 1  
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
 FT CARBOHYD 330 550  
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 8 8 Y -> N (IN PIR DATA BANK).  
 SQ SEQUENCE 550 AA; 61711 MW; 67BCD85F44736CFE CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFTENGEGMIDGNG 24  
 DB 330 GLFGAIGFTENGEGMIDGNG 352

RESULT 7  
 HEMA\_IADH7 STANDARD; PRT; 550 AA.  
 ID HEMA\_IADH7  
 AC P12568; O84018; O89470;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
 DE Hemagglutinin HA2 chain] (Fragment).  
 GN HA.  
 OS Influenza A virus (strain A/Duck/Hokkaido/10/85).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza virus A and B group; Influenza A viruses;  
 OC Influenza A virus.  
 OX NCBI\_TaxID=11363;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8725458; PubMed=2440178;  
 RA Kida H., Kawaoaka Y., Naeye C.W., Webster R.G.;  
 RT "Antigenic and genetic conservation of H3 influenza virus in wild  
 RT ducks.";  
 RT virology 159:109-119(1987).  
 RL -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.  
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M16743; AAA3149.1; -  
 DR PIR: G27813; HMIY15.

DR InterPro: IPR001364; Hemagglutn.  
 DR Pfam: PF00509; Hemagglutinin; 1.  
 DR ProDom: PD000225; Hemagglutn; 1.  
 KW Envelope protein; Hemagglutinin; Glycoprotein.  
 FT NON\_TER 1  
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
 FT CARBOHYD 330 550  
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 550 AA; 61761 MW; 6EF81793281D53EB CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFTENGEGMIDGNG 24  
 DB 330 GLFGAIGFTENGEGMIDGNG 352

RESULT 8  
 HEMA\_IADHK STANDARD; PRT; 550 AA.  
 ID HEMA\_IADHK  
 AC P43257;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
 DE Hemagglutinin HA2 chain] (Fragment).  
 GN HA.  
 OS Influenza A virus (strain A/Duck/Hong Kong/7/75).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza virus A and B group; Influenza A viruses;  
 OC Influenza A virus.  
 OX NCBI\_TaxID=11364;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91341491; PubMed=1875195;  
 RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;  
 RT "Molecular evidence for a role of domestic ducks in the introduction  
 RT of avian H3 influenza viruses to pigs in southern China, where the  
 RT A/Hong Kong/58 (H3N2) strain emerged.";  
 RT J. Gen. Virol. 72:2007-2010(1991).  
 RL -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.  
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: D00929; BAA00769.1; -  
 DR HSSP: P03437; SHMG.  
 DR InterPro: IPR001364; Hemagglutn.  
 DR Pfam: PF00509; Hemagglutinin; 1.  
 DR ProDom: PD000225; Hemagglutn; 1.  
 KW Envelope protein; Hemagglutinin; Glycoprotein.  
 FT NON\_TER 1  
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
 FT CARBOHYD 330 550  
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 550 AA; 61761 MW; 6EF81793281D53EB CRC64;

FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 550 AA: 61549 MW: 864639823FE1BA9 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDWYG 24  
DB 330 GLFGAIGFIENGWEGMIDWYG 352

RESULT 9  
HEMA\_IADHL STANDARD; PRT; 550 AA.  
AC P4326; 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;  
DE Hemagglutinin HA2 chain] (Fragment).  
GN HA.  
OS Influenza A virus (strain A/Duck/Hong Kong/54/76).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza A virus A and B group; Influenza A viruses;  
OC Influenza A virus.  
OX NCBI\_TaxID=45412;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91341491; PubMed=1875195;  
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;  
RT "Molecular evidence for a role of domestic ducks in the introduction  
RT of avian H3 influenza viruses to pigs in southern China, where the  
RT A/Hong Kong/68 (H3N2) strain emerged."  
RL J. Gen. Virol. 72:2007-2010(1991).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: D00931; BAA00771.1; -.  
DR HSSP: P03437; 3HMG.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR ProDom: PD000225; Hemagglutn; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON\_TER 1 1  
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 550 AA: 61718 MW: A351C56789E4BE9A CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 2.3e-07;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GLFGAIGFIENGWEGMIDWYG 24  
DB 330 GLFGAIGFIENGWEGMIDWYG 352

RESULT 10  
HEMA\_IAGHK STANDARD; PRT; 550 AA.  
AC P4326; 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;  
DE Hemagglutinin HA2 chain] (Fragment).  
GN HA.  
OS Influenza A virus (strain A/Goose/Hong Kong/10/76).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza A virus A and B group; Influenza A viruses;  
OC Influenza A virus.  
OX NCBI\_TaxID=45414;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91341491; PubMed=1875195;  
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;  
RT "Molecular evidence for a role of domestic ducks in the introduction  
RT of avian H3 influenza viruses to pigs in southern China, where the  
RT A/Hong Kong/68 (H3N2) strain emerged."  
RL J. Gen. Virol. 72:2007-2010(1991).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: D00930; BAA00770.1; -.  
DR HSSP: P03437; 3HMG.  
DR InterPro: IPR001364; Hemagglutn.  
DR Pfam: PF00509; Hemagglutinin; 1.  
DR ProDom: PD000225; Hemagglutn; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON\_TER 1 1  
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 550 AA: 61676 MW: 9A1B094DA28BACD2 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDWYG 24  
DB 330 GLFGAIGFIENGWEGMIDWYG 352

RESULT 11  
HEMA\_IADHL



```

ID HEMA_IJAZH2 STANDARD; PRT; 550 AA.
AC P1133; Q84019; Q84020;
DT 01-JUL-1989 (Rel. 11, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Swine/Hong Kong/81/78).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OX NCBI_TaxID=11497;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88101364; PubMed=3336940;
RA Kida H., Shortridge K.F., Webster R.G.;
RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China."
RL Virology 162:160-166(1988).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19057; AAA43212.1; -
DR PIR: B29971; HMTV53.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Hemagglutinin; Envelope protein; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 8 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 22 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61437 MW; 1F2A7E758C53ICE8 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGWY 24
DB 330 GLFGAIGFIENGWEGMIDGWY 352

RESULT 12
HEMA_IJAZH3 STANDARD; PRT; 550 AA.
ID HEMA_IJAZH3 STANDARD; PRT; 550 AA.
AC P1134; Q84025; Q84026;
DT 01-JUL-1989 (Rel. 11, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Swine/Hong Kong/126/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

```

```

OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11498;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88101364; PubMed=3336940;
RA Kida H., Shortridge K.F., Webster R.G.;
RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China."
RL Virology 162:160-166(1988).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19056; AAA43211.1; ALT_TERM.
DR PIR: A29971; HMTV52.
DR HSSP: P03437; 2HMG.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Hemagglutinin; Envelope protein; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 8 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 22 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61560 MW; 991F658BC02E24F2 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGWY 24
DB 330 GLFGAIGFIENGWEGMIDGWY 352

RESULT 13
HEMA_IJAZC STANDARD; PRT; 566 AA.
ID HEMA_IJAZC STANDARD; PRT; 566 AA.
AC P03437;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Alach/2/68).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OX NCBI_TaxID=150147;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=80254693; PubMed=7402351;
RA Verhoeven M., Fang R., Min Jou W., Devos R., Huybrecock D., Saman E., Fiers W.;
RT "Antigenic drift between the haemagglutinin of the Hong Kong

```



```
FT STRAND 310 311
FT STRAND 318 320
FT STRAND 323 324
FT STRAND 331 333
FT STRAND 337 337
FT STRAND 347 348
FT TURN 350 350
FT TURN 351 351
FT TURN 352 354
FT STRAND 355 355
FT STRAND 359 359
FT TURN 360 361
FT STRAND 367 372
FT TURN 374 375
FT STRAND 378 382
FT STRAND 383 401
FT STRAND 406 407
FT HELIX 421 474
FT STRAND 475 477
FT STRAND 482 485
FT HELIX 491 498
FT TURN 499 500
FT HELIX 505 515
SQ SEQUENCE 566 AA; 63415 MW; E395659C23CAFECA CRC64;
```

```
Query Match 55.1%; Score 134; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 2.4e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 2 GLFGAAGFIENGMEGMIDWYG 24
DB 346 GLFGAAGFIENGMEGMIDWYG 368
```

```
RESULT 14
HEMA_IAD3 STANDARD; PRT; 566 AA.
ID P26134;
AC 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;
Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Duck/Alberta/78/76).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11374;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114135; PubMed=1731092;
RA Bean W.J., Schell M., Katz J., Kawacka Y., Naeve C., Gorman O.,
RA Webster R.G.;
RT "Evolution of the H3 influenza virus hemagglutinin from human and
nonhuman hosts".
RL J. Virol. 66:1129-1138(1992).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL: M73771; -, NOT_ANNOTATED_CDS.
```

```
DR HSP; P03437; SHMG.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR Prodom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 344
FT CHAIN 346 566
FT CARBOHYD 23 23
FT CARBOHYD 24 24
FT CARBOHYD 38 38
FT CARBOHYD 54 54
FT CARBOHYD 181 181
FT CARBOHYD 301 301
FT CARBOHYD 499 499
SQ SEQUENCE 566 AA; 63534 MW; FE19AB6FF9415B89 CRC64;
```

```
Query Match 55.1%; Score 134; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 2.4e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 2 GLFGAAGFIENGMEGMIDWYG 24
DB 346 GLFGAAGFIENGMEGMIDWYG 368
```

```
RESULT 15
HEMA_IAD3 STANDARD; PRT; 566 AA.
ID P03442;
AC 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;
Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Duck/Ukraine/1/63).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11374;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8202542; PubMed=6169439;
RA Pang R., Min Jou W., Huyebroeck D., Devos R., Fiers W.;
RT "Complete structure of A/duck/Ukraine/63 influenza hemagglutinin
gene: animal virus as progenitor of human H3 Hong Kong 1968 influenza
hemagglutinin".
RL Cell 25:315-323(1981).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL: Y01087; CAA24271.1; -
DR PIR; A04053; HMTVDU.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR Prodom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
```

FT CHAIN 17 344 HEMAGGLUTININ HAI CHAIN.  
 FT CHAIN 346 566 HEMAGGLUTININ HAI CHAIN.  
 FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 566 AA; 63530 MW; E70F87F0AE1178F4 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 566;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-07;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDWTG 24  
 DB 346 GLFGAIGFIENGWEGMIDWTG 368

Search completed: July 1, 2002, 06:31:35  
 Job time: 672 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: July 1, 2002, 06:30:59 ; Search time 83.98 Seconds  
(without alignments)  
90.638 Million cell updates/sec

Title: US-09-461-684-5  
Perfect score: 243  
Sequence: 1 CGFAGAIAGTENGWGMID.....KKKKKKKKKKKKKKKKKK 44

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_19:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mmc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp rodent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*  
15: sp.yivirus:\*  
16: sp.bacterioplasmid:\*  
17: sp.archaeplastid:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	134	55.1	566	12 Q9DHG0	Q9dhg0 Influenza a
2	134	55.1	566	12 Q67125	Q67125 Influenza a
3	134	55.1	566	12 Q67126	Q67126 Influenza a
4	134	55.1	566	12 Q67132	Q67132 Influenza a
5	134	55.1	566	12 Q98052	Q98052 Influenza a
6	134	55.1	566	12 Q91M7	Q91M7 Influenza a
7	134	55.1	566	12 Q910M5	Q910M5 Influenza a
8	133	54.7	501	12 Q9DXK3	Q9dxk3 Influenza a
9	132	54.3	550	12 Q82499	Q82499 Influenza a
10	132	54.3	550	12 Q82753	Q82753 Influenza v
11	132	54.3	550	12 Q82498	Q82498 Influenza a
12	132	54.3	566	12 Q82496	Q82496 Influenza a
13	132	54.3	571	12 Q03909	Q03909 Influenza a
14	131	53.9	109	12 Q67050	Q67050 Influenza a
15	131	53.9	109	12 Q67051	Q67051 Influenza a
16	131	53.9	109	12 Q67052	Q67052 Influenza a

17	131	53.9	109	12 Q67053	Q67053 Influenza a
18	131	53.9	362	12 Q82513	Q82513 Influenza a
19	131	53.9	362	12 Q82517	Q82517 Influenza a
20	131	53.9	362	12 Q84174	Q84174 Influenza a
21	131	53.9	362	12 Q9GKD3	Q9gkd3 Influenza a
22	131	53.9	362	12 Q9GKD2	Q9gkd2 Influenza a
23	131	53.9	362	12 Q9GKD1	Q9gkd1 Influenza a
24	131	53.9	365	12 Q9DL25	Q9dl25 Influenza a
25	131	53.9	367	12 Q9DL22	Q9dl22 Influenza a
26	131	53.9	368	12 Q9DL29	Q9dl29 Influenza a
27	131	53.9	369	12 Q9DL26	Q9dl26 Influenza a
28	131	53.9	369	12 Q9DL06	Q9dl06 Influenza a
29	131	53.9	369	12 P87689	P87689 Influenza a
30	131	53.9	371	12 Q9DL24	Q9dl24 Influenza a
31	131	53.9	371	12 P87685	P87685 Influenza a
32	131	53.9	373	12 Q9DL20	Q9dl20 Influenza a
33	131	53.9	374	12 Q9DL21	Q9dl21 Influenza a
34	131	53.9	375	12 Q9DL27	Q9dl27 Influenza a
35	131	53.9	375	12 Q9DL05	Q9dl05 Influenza a
36	131	53.9	376	12 Q9DL30	Q9dl30 Influenza a
37	131	53.9	376	12 Q9DL04	Q9dl04 Influenza a
38	131	53.9	377	12 Q9E7P6	Q9e7p6 Influenza a
39	131	53.9	382	12 Q9DL03	Q9dl03 Influenza a
40	131	53.9	408	12 Q9E7P5	Q9e7p5 Influenza a
41	131	53.9	409	12 Q9G0L5	Q9g0l5 Influenza a
42	131	53.9	416	12 Q9GJC4	Q9gjc4 Influenza a
43	131	53.9	429	12 Q9G0L4	Q9g0l4 Influenza a
44	131	53.9	438	12 Q9G0L3	Q9g0l3 Influenza a
45	131	53.9	467	12 Q9E312	Q9e312 Influenza a

#### ALIGNMENTS

RESULT 1  
ID Q9DHG0 PRELIMINARY; PRT; 566 AA.  
AC Q9DHG0;  
DT 01-MAR-2001 (TREMURel. 16, Created)  
DT 01-MAR-2001 (TREMURel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMURel. 19, Last annotation update)  
DE HAEMAGGLUTININ PRECURSOR.  
OS Influenza A virus H3N2.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.  
OX NCBI\_taxid=41857;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CLONE 7A (H3N2);  
RA Mohsin M.A., Morris S.J., Smith H., Sweet C.;  
RT "Influenza virus-induced apoptosis: a dual role for viral neuraminidase".  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: HAEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HAEMAGGLUTININ FAMILY.  
CC EMBL; AJ289703; CAC18525.1; -.  
DR HSSP; P03437; 2V1U.  
DR InterPro; IPR001364; Hemagglutn.  
DR Pfam; PF00509; Hemagglutinin; 1.  
DR PRINTS; PR00329; HAEMAGGLUTN12.  
DR ProDom; PD000225; Hemagglutn; 1.  
KW Envelope protein; Glycoprotein; Hemagglutinin; Signal.  
FT SIGNAL 16  
FT SEQUENCE 566 AA; 63356 MW; 0BA601929300F72F CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFTENGWEGMIDGWC 24  
 DB 346 GLFGAIGFTENGWEGMIDGWC 368

## RESULT 2

O67125 PRELIMINARY: PRT: 566 AA.

AC O67125; 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DR HEMAGGLUTININ.

OS Influenza A virus.

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza virus A and B group; Influenza A viruses.

OX NCBI\_Taxid=11320;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-A/SEAL/MA/3911/92;  
 MEDLINE=95146951; PubMed=7844533;

RA Callan R.J., Early G., Kida H., Hinshaw V.S.;  
 RT "The appearance of H3 influenza viruses in seals."

RL J. Gen. Virol. 76:199-203(1995).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

DR EMBL: L32024; AAA64228.1; -

DR HSSP: P03437; 2V10.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR PRINTS: PR00329; HEMAGGLUTIN12.

DR ProDom: PD000225; Hemagglutn; 1.

KW Envelope protein; Glycoprotein; Hemagglutinin.

SO SEQUENCE 566 AA; 63456 MW; AE56302A9EB89F CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;

Best Local Similarity 100.0%; Pred. No. 4e-08;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFTENGWEGMIDGWC 24

DB 346 GLFGAIGFTENGWEGMIDGWC 368

## RESULT 3

O67126 PRELIMINARY: PRT: 566 AA.

AC O67126; 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DR HEMAGGLUTININ.

OS Influenza A virus.

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza virus A and B group; Influenza A viruses.

OX NCBI\_Taxid=11320;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-A/SEAL/MA/3984/92;  
 MEDLINE=95146951; PubMed=7844533;

RA Callan R.J., Early G., Kida H., Hinshaw V.S.;  
 RT "The appearance of H3 influenza viruses in seals."

RL J. Gen. Virol. 76:199-203(1995).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
 DR EMBL: L32024; AAA64228.1; -  
 DR HSSP: P03437; 2V10.  
 DR InterPro: IPR001364; Hemagglutn.  
 DR Pfam: PF00509; Hemagglutinin; 1.  
 DR PRINTS: PR00329; HEMAGGLUTIN12.  
 DR ProDom: PD000225; Hemagglutn; 1.  
 KW Envelope protein; Glycoprotein; Hemagglutinin.  
 SO SEQUENCE 566 AA; 63441 MW; 590576C4CEB7D08 CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;

Best Local Similarity 100.0%; Pred. No. 4e-08;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFTENGWEGMIDGWC 24

DB 346 GLFGAIGFTENGWEGMIDGWC 368

## RESULT 4

O67132 PRELIMINARY: PRT: 566 AA.

AC O67132; 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DR HEMAGGLUTININ.

OS Influenza A virus (strain A/Alchi/2/68).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza virus A and B group; Influenza A viruses.

OX NCBI\_Taxid=150147;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-A/Alchi/2/68;

RA Min J.W., Verhoeven M., Fang R.-X., Devos R., Huybrecock D.,  
 RA Fiers W.;

RT "Shift and drift in influenza viruses."

RL (in) Carille M.J., Collins J.F., Moseley B.E. B. (eds.);  
 RL SYMPOSIUM OF THE SOCIETY FOR GENERAL MICROBIOLOGY, pp.285-311,  
 RL Cambridge University Press, New York (1981).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

DR EMBL: M55059; AAA43239.1; -

DR HSSP: P03437; 1HGE.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR PRINTS: PR00329; HEMAGGLUTIN12.

DR ProDom: PD000225; Hemagglutn; 1.

KW Envelope protein; Glycoprotein; Hemagglutinin.

FT CHAIN 1 344 HEMAGGLUTININ.  
 FT CHAIN 346 566 NEURAMINIDASE.

SO SEQUENCE 566 AA; 63441 MW; ESD1B97DB96FECA CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;

Best Local Similarity 100.0%; Pred. No. 4e-08;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFTENGWEGMIDGWC 24

DB 346 GLFGAIGFTENGWEGMIDGWC 368

## RESULT 5

O68052 PRELIMINARY: PRT: 566 AA.

AC O68052; 01-FEB-1997 (Tremblrel. 02, Created)

DT 01-FEB-1997 (TREMblrel. 02, last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)  
 DE HEMAGGLUTININ PRECURSOR (FRAGMENT).  
 OS Influenza A virus.  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 CC Influenza virus A and B group; Influenza A viruses.  
 OX NCBI\_TaxID=11320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=81053698; PubMed=6253883;  
 RX Both G.W., Sleight M.J.;  
 RT "Complete nucleotide sequence of the haemagglutinin gene from a human  
 RT Influenza virus of the Hong Kong subtype."  
 RL Nucleic Acids Res. 8:2561-2575(1980).  
 RN [2]  
 RP SEQUENCE OF 17-344 FROM N.A.  
 RA MEDLINE=81194918; PubMed=6164798;  
 RX Sleight M.J., Both G.W., Underwood P.A., Bender V.J.;  
 RT "Antigenic drift in the hemagglutinin of the Hong Kong influenza  
 RT subtype: Correlation of amino acid changes with alterations in viral  
 RT antigenicity."  
 RL J. Virol. 37:845-853(1981).  
 RN [3]  
 RP SEQUENCE OF 17-566 FROM N.A.  
 RA MEDLINE=82033276; PubMed=6169843;  
 RX Both G.W., Sleight M.J.;  
 RT "Conservation and variation in the hemagglutinins of Hong Kong subtype  
 RT Influenza viruses during antigenic drift."  
 RL J. Virol. 39:845-853(1981).  
 RN [4]  
 RP CELL RECEPTORS AND FOR INITIATING THE VIRUS TO  
 RN [5]  
 RP SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
 RN [6]  
 RP (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
 RN [7]  
 RP -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
 RN [8]  
 RP EMBL: J02135; AAA3189.1; -  
 RN [9]  
 RP HSP: P03437; 1HG.  
 RN [10]  
 RP InterPro: IPR001364; Hemagglutn.  
 RN [11]  
 RP Pfam: PF00509; Hemagglutinin; 1.  
 RN [12]  
 RP PRINTS: PR00329; HEMAGGLUTIN12.  
 RN [13]  
 RP ProDom: PD000225; Hemagglutn; 1.  
 RN [14]  
 RP Envelope protein; Glycoprotein; Hemagglutinin; Signal.  
 RN [15]  
 RP SIGNAL 1 16 POTENTIAL.  
 RN [16]  
 RP CHAIN 17 344 POTENTIAL.  
 RN [17]  
 RP CHAIN 346 566 POTENTIAL.  
 RN [18]  
 RP SEQUENCE 566 AA; 63414 MW; C447FD465B54FCF9 CRC64;  
 SQ  
 Query Match 55.1%; Score 134; DB 12; Length 566;  
 Best Local Similarity 100.0%; Pred. No. 4e-08;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GLFGAIGFIENGEGMIDGMYG 24  
 DB 346 GLFGAIGFIENGEGMIDGMYG 368  
 RESULT 6  
 ID Q91MA7 PRELIMINARY; PRT; 566 AA.  
 AC Q91MA7  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)  
 DE HEMAGGLUTININ.  
 OS Influenza A virus (A/Hong Kong/1/68(H3N2)).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 CC Influenza virus A and B group; Influenza A viruses.  
 OX NCBI\_TaxID=108859;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-A/HONG KONG/1/68(H3N2);  
 RX MEDLINE=21287244; PubMed=11371620;  
 RT Brown E.G., Liu H., Kit L.C., Baird S., Nesrallah M.;

RT "Pattern of mutation in the genome of influenza A virus on adaptation  
 RT to increased virulence in the mouse lung: Identification of functional  
 RT themes."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).  
 DR EMBL: AF348176; AAK51718.1; -  
 SQ SEQUENCE 566 AA; 63387 MW; 01BB0D465BE158E1 CRC64;  
 QY 2 GLFGAIGFIENGEGMIDGMYG 24  
 DB 346 GLFGAIGFIENGEGMIDGMYG 368  
 RESULT 7  
 ID Q910M5 PRELIMINARY; PRT; 566 AA.  
 AC Q910M5  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, last sequence update)  
 DE HEMAGGLUTININ.  
 OS Influenza A virus (A/Hong Kong/1/68(H3N2)).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 CC Influenza virus A and B group; Influenza A viruses.  
 OX NCBI\_TaxID=108859;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-A/HONG KONG/1/68(H3N2);  
 RX MEDLINE=21287244; PubMed=11371620;  
 RT Brown E.G., Liu H., Kit L.C., Baird S., Nesrallah M.;  
 RT "Pattern of mutation in the genome of influenza A virus on adaptation  
 RT to increased virulence in the mouse lung: Identification of functional  
 RT themes."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).  
 DR EMBL: AF348179; AAK51721.1; -  
 DR EMBL: AF348177; AAK51720.1; -  
 DR EMBL: AF348178; AAK51720.1; -  
 SQ SEQUENCE 566 AA; 63530 MW; 7CB9F5BAF1E6E9F4 CRC64;  
 QY 2 GLFGAIGFIENGEGMIDGMYG 24  
 DB 346 GLFGAIGFIENGEGMIDGMYG 368  
 Query Match 55.1%; Score 134; DB 12; Length 566;  
 Best Local Similarity 100.0%; Pred. No. 4e-08;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GLFGAIGFIENGEGMIDGMYG 24  
 DB 346 GLFGAIGFIENGEGMIDGMYG 368  
 RESULT 8  
 ID Q9DXE3 PRELIMINARY; PRT; 301 AA.  
 AC Q9DXE3  
 DT 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)  
 DE HEMAGGLUTININ (FRAGMENT).  
 GN HA1.  
 OS Influenza A virus (A/Shorebird/Taiwan/31-4/99).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 CC Influenza virus A and B group; Influenza A viruses.  
 OX NCBI\_TaxID=140665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-A/SHOREBIRD/TAIWAN/31-4/99;  
 RC Lee M.S., Cheng P.C., Shien J.H., Cheng M.C., Lee L.H., Shieh H.K.,  
 RT "Identification and subtyping of avian influenza virus by reverse  
 RT transcription-polymerase chain reaction."  
 RT Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
 DR EMBL: AF311750; AAG3016.1; -.  
 DR InterPro: IPR001364; Hemagglutinin.  
 DR PRINTS: PR003329; Hemagglutinin.  
 DR ProDom: PD000225; Hemagglutn; 1.  
 KW Envelope protein; Glycoprotein; Hemagglutinin.  
 FT NON\_TER 1 301  
 FT SEQUENCE 301 AA; 32701 MW; 62A403758B764D57 CRC64;

Query Match 54.7%; Score 133; DB 12; Length 301;  
 Best Local Similarity 95.7%; Pred. No. 2.7e-08;  
 Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GFGAIAFGIENGMEGMDGWTG 24  
 |||||  
 DB 250 GFGAIAFGIENGMEGMDGWTG 272

RESULT 9  
 O82499 PRELIMINARY; PRT; 550 AA.  
 AC O82499;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE HEMAGGLUTININS HA1 AND HA2 (FRAGMENT).  
 OS Influenza A virus.  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza virus A and B group; Influenza A viruses.  
 OK NCBI\_TaxID=11320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A/PHILIPPINES/2/82/BS (H3N2);  
 RA Hartley C.A., Ward A.C., Anders E.M.;  
 RT "Virulence of influenza A virus for mice is associated with loss of  
 RT oligosaccharide from the hemagglutinin molecule."  
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
 DR EMBL: J06889; AA18782.1; -.  
 DR HSSP: P03437; 2V1U;  
 DR InterPro: IPR001364; Hemagglutn.  
 DR Pfam: PF00509; Hemagglutinin; 1.  
 DR PRINTS: PR00329; Hemagglutn12.  
 DR ProDom: PD000225; Hemagglutn; 1.  
 KW Envelope protein; Glycoprotein; Hemagglutinin.  
 FT NON\_TER 1 328  
 FT CHAIN 1 330 HA1.  
 FT CHAIN 330 550 HA2.  
 FT SEQUENCE 550 AA; 61772 MW; 50BD62B6FE11FD8 CRC64;

Query Match 54.3%; Score 132; DB 12; Length 550;  
 Best Local Similarity 95.7%; Pred. No. 6.7e-08;  
 Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GFGAIAFGIENGMEGMDGWTG 24  
 |||||  
 DB 330 GFGAIAFGIENGMEGMDGWTG 352

RESULT 10  
 O82753 PRELIMINARY; PRT; 550 AA.  
 ID O82753

AC O82753;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE HEMAGGLUTININ (FRAGMENT).  
 OS Influenza virus.  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Unclassified Orthomyxoviridae.  
 OK NCBI\_TaxID=11305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A/PHILIPPINES/2/82/BS/ML10 (H3N2);  
 RX MEDLINE=97300854; Pubmed=9155874;  
 RA Hartley C.A., Reading P.C., Ward A.C., Anders E.M.;  
 RT "Changes in the hemagglutinin molecule of influenza type A (H3N2)  
 RT virus associated with increased virulence for mice."  
 RL Arch. Virol. 142:75-86(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A/PHILIPPINES/2/82/BS/ML10 (H3N2);  
 RX MEDLINE=97456249; Pubmed=9311563;  
 RA Ward A.C.;  
 RT "Virulence of influenza A virus for mouse lung."  
 RL Virus Genes 14:187-194(1997).  
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
 DR EMBL: U06805; AAC79579.1; -.  
 DR HSSP: P03437; 2V1U;  
 DR InterPro: IPR001364; Hemagglutn.  
 DR Pfam: PF00509; Hemagglutinin; 1.  
 DR PRINTS: PR00329; Hemagglutn12.  
 DR ProDom: PD000225; Hemagglutn; 1.  
 KW Envelope protein; Glycoprotein; Hemagglutinin.  
 FT NON\_TER 1 328  
 FT CHAIN 1 330 HAEMAGGLUTININ HA1.  
 FT CHAIN 330 550 HAEMAGGLUTININ HA2.  
 FT SEQUENCE 550 AA; 61745 MW; 692A49DE678AC4BC CRC64;

Query Match 54.3%; Score 132; DB 12; Length 550;  
 Best Local Similarity 95.7%; Pred. No. 6.7e-08;  
 Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GFGAIAFGIENGMEGMDGWTG 24  
 |||||  
 DB 330 GFGAIAFGIENGMEGMDGWTG 352

RESULT 11  
 O82498 PRELIMINARY; PRT; 550 AA.  
 AC O82498;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE HEMAGGLUTININS HA1 AND HA2 (FRAGMENT).  
 OS Influenza A virus.  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza virus A and B group; Influenza A viruses.  
 OK NCBI\_TaxID=11320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A/PHILIPPINES/2/82 (H3N2);  
 RA Hartley C.A., Ward A.C., Anders E.M.;  
 RT "Virulence of influenza virus for mice is associated with loss of  
 RT oligosaccharide from the hemagglutinin molecule."  
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A/PHILIPPINES/2/82 (H3N2);



RX MEDLINE=8185444; PubMed=3356226;  
 RA Nakajima S., Takeuchi Y., Nakajima K.;  
 RT "Location on the evolutionary tree of influenza H3 haemagglutinin  
 RT genes of Japanese strains isolated during the 1985-6 season.";  
 RL Epidemiol. Infect. 100:301-310(1988).  
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
 DR EMBL: U08856; AAA18781.1; -.  
 DR HSSP: P03437; 2YIU.  
 DR InterPro: IPR001364; Hemagglutn.  
 DR Pfam: PF00509; Hemagglutinin; 1.  
 DR PRINTS: PR00329; HEMAGGLUTIN12.  
 DR ProDom: PD000225; Hemagglutn; 1.  
 DR Envelope protein; Glycoprotein; Hemagglutinin.  
 FT NON\_TER 1 1  
 FT CHAIN 1 328 HA1.  
 FT CHAIN 330 350 HA2.  
 SQ SEQUENCE 550 AA; 61802 MW; 1144131CE5A1F6A CRC64;

Query Match 54.3%; Score 132; DB 12; Length 550;  
 Best Local Similarity 95.7%; Pred. No. 6.7e-08;  
 Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLEFAGIAGFIENGEGMIDGMYG 24  
 1:|||||  
 DB 330 GLEFAGIAGFIENGEGMIDGMYG 352

RESULT 12  
 Q82496 PRELIMINARY; PRT; 566 AA.  
 AC Q82496;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HEMAGGLUTININ.  
 GN HA.  
 OS Influenza A virus.  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza virus A and B group; Influenza A viruses.  
 OC NCBI\_Taxid=11320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A/SWINE/ANGE-GARDIEN/150/90(H3N2);  
 RX MEDLINE=95205091; PubMed=7897358;  
 RA Brikour M.H., Frost E.H., Deslandes S., Talbot B., Weber J.M.,  
 RA Elazhary Y.;  
 RT "Recent H3N2 swine influenza virus with haemagglutinin and  
 RT nucleoprotein genes similar to 1975 human strains.";  
 RL J. Gen. Virol. 76:697-703(1995).  
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
 DR EMBL: U07146; AAA85781.1; -.  
 DR HSSP: P03437; 2YIU.  
 DR InterPro: IPR001364; Hemagglutn.  
 DR Pfam: PF00509; Hemagglutinin; 1.  
 DR PRINTS: PR00329; HEMAGGLUTIN12.  
 DR ProDom: PD000225; Hemagglutn; 1.  
 DR Envelope protein; Glycoprotein; Hemagglutinin.  
 SQ SEQUENCE 566 AA; 63362 MW; 1FB8319A561E2FF CRC64;

Query Match 54.3%; Score 132; DB 12; Length 566;  
 Best Local Similarity 95.7%; Pred. No. 6.9e-08;  
 Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLEFAGIAGFIENGEGMIDGMYG 24  
 1:|||||  
 DB 346 GLEFAGIAGFIENGEGMIDGMYG 368

RESULT 13  
 Q03909 PRELIMINARY; PRT; 571 AA.  
 ID Q03909;  
 AC Q03909;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HEMAGGLUTININ PRECURSOR.  
 GN HA.  
 OS Influenza A virus.  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza virus A and B group; Influenza A viruses.  
 OC NCBI\_Taxid=11320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Guo Y., Wang M., Kawachi Y., Gorman O.T., Ito T., Webster R.G.;  
 RL Submitted (XXX-1991) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.  
 DR EMBL: M65018; AAA43151.1; -.  
 DR HSSP: P03437; 2YIU.  
 DR InterPro: IPR001364; Hemagglutn.  
 DR Pfam: PF00509; Hemagglutinin; 1.  
 DR PRINTS: PR00329; HEMAGGLUTIN12.  
 DR ProDom: PD000225; Hemagglutn; 1.  
 DR Envelope protein; Hemagglutinin; Glycoprotein; Signal.  
 FT SIGNAL 1 21  
 FT CHAIN 22 349 HA1 CHAIN.  
 FT CHAIN 350 571 HA2 CHAIN.  
 FT DISULFID 36 487 BY SIMILARITY.  
 FT DISULFID 73 298 BY SIMILARITY.  
 FT DISULFID 85 97 BY SIMILARITY.  
 FT DISULFID 160 494 BY SIMILARITY.  
 SQ SEQUENCE 571 AA; 64104 MW; 718DAA0F291CE349 CRC64;

Query Match 54.3%; Score 132; DB 12; Length 571;  
 Best Local Similarity 95.7%; Pred. No. 7e-08;  
 Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLEFAGIAGFIENGEGMIDGMYG 24  
 1:|||||  
 DB 351 GLEFAGIAGFIENGEGMIDGMYG 373

RESULT 14  
 Q67050 PRELIMINARY; PRT; 109 AA.  
 ID Q67050;  
 AC Q67050;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HEMAGGLUTININ (FRAGMENT).  
 GN HA.  
 OS Influenza A virus.  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza virus A and B group; Influenza A viruses.  
 OC NCBI\_Taxid=11320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A/FUKUOKA/C29/85 (H3N2);  
 RX MEDLINE=81030832; PubMed=7421990;  
 RA Gething M.-J., Bye J., Skehel J., Waterfield M.;  
 RT "Cloning and dna sequence of double-stranded copies of haemagglutinin  
 RT genes from h2 and h3 strains elucidates antigenic shift and drift in  
 RT human influenza virus.";  
 RL Nature 287:301-306(1980).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=A/FUKUOKA/C29/85 (H3N2);  
 RA MEDLINE=93333219; PubMed=7682624;  
 RX Okuno Y., Isegawa Y., Sasao F., Ueda S.;  
 RT "A common neutralizing epitope conserved between the hemagglutinins of  
 influenza A virus H1 and H2 strains.";  
 RL J. Virol. 67:2552-2558(1993).  
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
 CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
 (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
 DR EMBL: D13581; BAA02776.1; -.  
 DR HSSP: P03437; IRTV.  
 DR InterPro: IPR001364; Hemagglutn.  
 DR Pfam: PF00509; Hemagglutinin; 1.  
 DR PRINTS: PR00329; HEMAGGLUTIN12.  
 DR ProDom: PD000225; Hemagglutn; 1.  
 KW Envelope protein; Glycoprotein; Hemagglutinin.  
 FT NON\_TER 1  
 FT 109  
 SO SEQUENCE 109 AA; 12305 MW; 17EC66753C48672F CRC64;

Query Match 53.9%; Score 131; DB 12; Length 109;  
 Best Local Similarity 91.3%; Pred. No. 1.5e-08;  
 Matches 21; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIENGWEGMIDGWG 24  
 ID 40 GIFGAIAGFIENGWEGMIDGWG 62

RESULT 15  
 ID 067051 PRELIMINARY; PRT; 109 AA.  
 AC 067051;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE HEMAGGLUTININ (FRAGMENT).  
 GN HA.  
 OS Influenza A virus.  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza virus A and B group; Influenza A viruses.  
 OX NCBI\_TaxID=11320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A/SICHUAN/2/87 (H3N2);  
 RX MEDLINE=81030852; PubMed=7421990;  
 RA Gething M.-J., Bye J., Skehel J., Waterfield M.;  
 RT "Cloning and dna sequence of double-stranded copies of haemagglutinin  
 genes from h2 and h3 strains elucidates antigenic shift and drift in  
 human influenza virus.";  
 RT Nature 287:301-306(1980).  
 RL [2]  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A/SICHUAN/2/87 (H3N2);  
 RX MEDLINE=93233219; PubMed=7682624;  
 RA Okuno Y., Isegawa Y., Sasao F., Ueda S.;  
 RT "A common neutralizing epitope conserved between the hemagglutinins of  
 influenza A virus H1 and H2 strains.";  
 RL J. Virol. 67:2552-2558(1993).  
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
 CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
 (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
 DR EMBL: D13582; BAA02777.1; -.  
 DR HSSP: P03437; IHGE.  
 DR InterPro: IPR001364; Hemagglutn.  
 DR Pfam: PF00509; Hemagglutinin; 1.  
 DR PRINTS: PR00329; HEMAGGLUTIN12.

DR ProDom: PD000225; Hemagglutn; 1.  
 KW Envelope protein; Glycoprotein; Hemagglutinin.  
 FT NON\_TER 1  
 FT 109  
 SO SEQUENCE 109 AA; 12293 MW; 17EC66752DB8672F CRC64;

Query Match 53.9%; Score 131; DB 12; Length 109;  
 Best Local Similarity 91.3%; Pred. No. 1.5e-08;  
 Matches 21; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIENGWEGMIDGWG 24  
 ID 40 GIFGAIAGFIENGWEGMIDGWG 62

Search completed: July 1, 2002, 06:30:59  
 Job time: 691 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:19:25 ; Search time 98.97 Seconds

(without alignments)  
49,381 Million cell updates/sec

Title: US-09-461-684-5  
Perfect score: 243  
Sequence: 1 CGLFGAIGFTENGEGMID.....KKKKKKKKKKKKKKKKKK 44

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: A\_Geneseq\_032802:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	243	100.0	44	21	AA13784
2	192	79.0	40	22	AA151431
3	143	58.8	24	21	AA13782
4	143	58.8	26	13	AA26004
5	143	58.8	26	18	AA134269
6	143	58.8	26	18	AA134269
7	143	58.8	26	18	AA134269
8	143	58.8	26	18	AA134269
9	143	58.8	26	18	AA134269
10	143	58.8	26	18	AA134269
11	143	58.8	26	18	AA134269

12	134	55.1	28	17	AA198024	Fusogenic peptide
13	134	55.1	30	21	AA198680	Core polypeptide f
14	134	55.1	30	22	AA198088	Viral Pp178/107-11
15	134	55.1	30	22	AA198257	Viral core polypep
16	134	55.1	30	22	AA198364	Pp178-11ke/Dp107-1
17	134	55.1	30	22	AA198081	Core polypeptide T
18	134	55.1	175	15	AA198579	Sequence of SHA2.
19	134	55.1	223	22	AA194032	Peptide sequence f
20	134	55.1	347	15	AA198391	Stem region of A2/
21	134	55.1	566	15	AA198390	Full length H3N2 i
22	134	55.1	685	22	AA198783	Modified clostridi
23	132	54.3	29	20	AA198710	Influenza virus an
24	132	54.3	221	14	AA198710	Sequence of the HA
25	132	54.3	221	14	AA198710	Sequence of the HA
26	132	54.3	221	14	AA198710	Sequence of the HA
27	132	54.3	221	14	AA198710	Sequence of the HA
28	132	54.3	221	14	AA198710	Sequence of the HA
29	132	54.3	221	14	AA198710	Sequence of the HA
30	132	54.3	221	14	AA198710	Sequence of the HA
31	131	53.9	365	8	AA198710	Immunogenic fragme
32	131	53.9	365	8	AA198710	Sequence of the X-
33	131	53.9	365	8	AA198710	Equine influenza v
34	131	53.9	365	8	AA198710	Equine hemagglutin
35	131	53.9	365	8	AA198710	Equine hemagglutin
36	131	53.9	365	8	AA198710	Equine hemagglutin
37	131	53.9	365	8	AA198710	Equine hemagglutin
38	131	53.9	365	8	AA198710	Equine hemagglutin
39	131	53.9	365	8	AA198710	Equine hemagglutin
40	131	53.9	365	8	AA198710	Equine hemagglutin
41	131	53.9	365	8	AA198710	Equine hemagglutin
42	131	53.9	365	8	AA198710	Equine hemagglutin
43	131	53.9	365	8	AA198710	Equine hemagglutin
44	131	53.9	365	8	AA198710	Equine hemagglutin
45	131	53.9	365	8	AA198710	Equine hemagglutin

## ALIGNMENTS

RESULT 1	AA13784	standard; peptide: 44 AA.
ID	AA13784	
XX	AA13784	
AC	AA13784	
XX	AA13784	
DT	10-NOV-2000	(first entry)
XX	AA13784	
DE	Soluble tandem HA/ PK peptide conjugate.	
XX	AA13784	
KW	pk peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;	
KW	major histocompatibility complex class I; MHC class I; antigen; tumour;	
KW	prostate; breast; multiple myeloma; HA peptide.	
XX	AA13784	
OS	Unidentified.	
XX	AA13784	
PN	WO200035949-A1.	
XX	AA13784	
PD	22-JUN-2000.	
XX	AA13784	
PF	14-DEC-1999;	99WO-US29724.
XX	AA13784	
PR	14-DEC-1998;	98US-0112324.
XX	AA13784	
PA	(DEND-) DENDRON CORP.	
XX	AA13784	
PI	Laus R, Hakim I, Vidovic D;	
XX	AA13784	
DR	WPI; 2000-442365/38.	
XX	AA13784	
PT	Antigens modified by the covalent addition of a peptide that	
PT	facilitates entry into antigen presenting cells, useful for producing	
PT	compositions for immunizing against tumors and pathogens -	



```

RESULT 4
AAAR26004
ID AAR26004 standard; peptide; 26 AA.
AC AAR26004;
DT 26-JAN-1993 (first entry)
XX Influenza fusion peptide #1.
DE Influenza fusion peptide #1.
XX Haemagglutinin; cysteine; liposome.
XX Synthetic.
OS EP497997-A.
PN EP497997-A.
XX 12-AUG-1992.
PD 12-AUG-1992.
XX 02-FEB-1991; 91EP-0101414.
PF 02-FEB-1991; 91EP-0101414.
XX 02-FEB-1991; 91EP-0101414.
PR 02-FEB-1991; 91EP-0101414.
XX (NIKA-) NIKA HEALTH PROD LTD.
PA (Gluck R, Herrmann P, Klein P;
XX WPI; 1992-270078/33.
DR WPI; 1992-270078/33.
XX Drug-contg. phospholipid bilayer vesicle with cell-specific
PT markers on membrane - where markers have at least 90 per cent
PT biological activity, used as pharmaceuticals against e.g. cancer
PT and AIDS
XX Disclosure; Fig 2; 13pp; English.
PS The sequences given in AAR26004-19 are influenza fusion peptides,
XX derived from the influenza virus haemagglutinin gene which were
CC used for the preparation of synthetic membrane vesicles. The
CC arrangement of at least one, pref. three cysteine residues at one
CC end of these peptides has been found useful for the fusion activity,
CC for the fusion of the liposome to the target cell membrane. The
CC liposomes produced using these peptides can contain at least one
CC active drug and can be used to target cells infected with cancer
CC or AIDS.
XX Sequence 26 AA;
SQ
Query Match 58.8%; Score 143; DB 13; Length 26;
Best Local Similarity 100.0%; Pred. No. 7e-11;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY 1 CGLFGAIGFIENGWEGMIDGWYG 24
Db |||||||
3 cglfgaigfiengwegmidgwgy 26
RESULT 5
AAAR34269
ID AAW34269 standard; peptide; 26 AA.
AC AAW34269;
DT 14-MAY-1998 (first entry)
XX Synthetic lipid vesicle fusion peptide 1.
DE Virosome; vesicle; cationic lipid bilayer; viral fusion peptide;
XX drug delivery system; membrane; gene therapy; diagnosis; treatment;
KW cancer; leukaemia; viral infection.
XX Synthetic.
XS

```

```

XX XX MW0741834-A1.
PD 13-NOV-1997.
XX XX
PF 04-MAY-1997; 97MO-EP02268.
FR 08-MAY-1996; 96EP-0107282.
XX XX
PA (NIKA-) NIKA HEALTH PROD LTD.
XX XX
PI Glueck R, Klein P, Maelit ER;
DR WPI: 1997-558673/51.
XX XX
PT Vesicle - used for delivery of genetic material to cells, especially
PT peptide - used for delivery of genetic material to cells, especially
PT for gene therapy of cancer, leukemia and viral infections
XX XX
PS Disclosure: Page 10; 52pp; English.
XX XX
CC Peptides AAM34269-WJ4284 represent novel lipid vesicles with positively
CC charged lipid bilayer membranes composed of a cationic and/or
CC polycationic lipid and at least one natural or synthetic viral fusion
CC peptide integrated in, or covalently linked to, the membrane. Such
CC peptides are used as drug delivery systems, preferably for
CC (non-)specific delivery of genetic material to target cells or tissues,
CC particularly for diagnosis, treatment (especially antisense treatment)
CC of cancer, leukaemia and viral infections in humans or animals. Genetic
CC material is delivered, without infection, to resting or proliferating
CC cells, in vitro or in vivo. When the genetic material is an antisense
CC molecule, it is targeted to mRNA encoding a (proto)oncogene. The
CC continuous lipid layer does not leak. The peptides do not need to fuse
CC with, or destabilise, plasma membranes in order to enter the cytoplasm,
CC since the fusion peptide ensures cell penetration by endocytosis (after
CC which fusion of the vesicle and endosomal membrane occurs). The genetic
CC material thus has a greater chance of reaching the nucleus before it is
CC degraded or expelled. Transfer of the material is 1000-20000 times more
CC efficient than when using liposomes or conventional virosomes, so
CC smaller doses can be used, avoiding possible toxicity associated with the
CC genetic material.
XX XX
SQ Sequence 26 AA:
OY Query Match 58.8%; Score 143; DB 18; Length 26;
Best Local Similarity 100.0%; Pred. No. 7e-11;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 CGLFGAINGTENGWEGMIDGWYG 24
| | | | | | | | | | | | | | | | | |
3 cglfgaagflengwgmldgwyg 26
DE RESULT 6
ID AAB70090
XX AAB70090 standard; peptide: 26 AA.
XX AC AAB70090:
XX 14-MAY-2001 (first entry)
XX DE Cationic virosome crosslinker #1 for polypeptide binding.
XX XX Cationic virosome; cytostatic; gene therapy; lipid bilayer vesicle;
XX KM viral haemagglutinin; drug delivery; cancer; leukaemia; viral infection;
XX crosslinker.
XX OS Unidentified.
XX PN NZ504444-A.
XX PD 24-NOV-2000.
XX XX

```

PF 10-MAY-2000; 2000NZ-0504444.  
XX  
XX  
PR 10-MAY-2000; 2000NZ-0332666.  
XX  
XX  
PA (NIKA-) NIKA HEALTH PROD LTD.  
XX  
XX  
PI Waliti ER, Gluck R, Klein P;  
XX  
XX  
DR WPI; 2001-233042/24.  
XX  
XX  
PT Lipid bilayer vesicle, useful for delivering drugs to target cells such  
PT as cancer, leukemic, or virally infected cells, comprises viral  
PT glycoproteins in positively charged membrane  
XX  
XX  
PS Disclosure; Page 14; 41pp; English.  
XX  
XX  
CC The present sequence is a crosslinker which is capable of linking to the  
CC surface of a novel cationic virosome and is capable of binding  
CC polypeptides. The cationic virosome is a lipid bilayer vesicle comprising  
CC a membrane with a net positive charge and contains 5-30 weight % based on  
CC total lipids, 1,3-dioleoyloxy-2-(6-carboxy-spermyl)-propyl-amine  
CC (DOSPER) together with other lipids and at least one active fusogenic  
CC peptide. The fusogenic peptide is a viral haemagglutinin that causes the  
CC vesicle to be internalised by target cells through phagocytosis or  
CC endocytosis. The virosome is useful for delivering a desired drug or  
CC substance, preferably a nucleic acid, to target cells (resting or  
CC proliferating mammalian cells) such as cancer cells, leukaemic cells or  
CC virally infected cells in vitro, in diagnostic or medical applications  
CC and for the manufacture of medicament for carrying out non-infectious  
CC gene therapy.  
XX  
XX  
SQ Sequence 26 AA:  
  
Query Match 58.8%; Score 143; DB 22; Length 26;  
Best Local Similarity 100.0%; Pred. No. 7e-11;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GGLFGAIGFTENGEGMIDGWY 24  
Db |||||  
3 cglfgalagftengwegmidgwy 26  
  
RESULT 7  
AAW29226  
ID AAW29226 standard; peptide; 23 AA.  
XX  
XX  
AC AAW29226;  
XX  
XX  
DT 02-MAR-1998 (first entry)  
XX  
XX  
DE Membrane active acidic peptide INF6 from WO9730170.  
XX  
XX  
KW Membrane active; acidic peptide; cationic lipid; transfection;  
KW gene therapy; tumour; vaccine; cytokine gene.  
XX  
XX  
OS Synthetic.  
OS Influenza virus.  
XX  
XX  
FN WO9730170-A1.  
XX  
XX  
PD 21-AUG-1997.  
XX  
XX  
PF 13-FEB-1997; 97WO-EP00649.  
XX  
XX  
PR 15-FEB-1996; 96DE-4005548.  
XX  
XX  
PA (BOEH) BOEHRINGER INGELHEIM INT GMBH.  
XX  
XX  
PI Kichler A, Mechtler K, Wagner E;  
XX  
XX  
DR WPI; 1997-425044/39.  
XX  
XX  
PS

PT Composition containing nucleic acid complexed with cationic lipid  
PT and a membrane-active, acidic peptide - used to increase efficiency  
PT of transfection by destabilising the endosomal membrane, especially  
PT useful for gene therapy  
XX  
XX  
PS Claim 4; Page 33; 63pp; German.  
XX  
XX  
CC A new composition has been developed for transfecting higher eukaryotic  
CC cells contains: (i) a complex of nucleic acid to be expressed plus a  
CC suboptimal concentration of one or more cationic lipids, and optionally  
CC at least one auxiliary lipid; and (ii) the new feature, at least one  
CC membrane-active, acidic peptide; where the ratio total positive  
CC charges:total negative charges in the composition is about 0-3,  
CC preferably 0-2. The present sequence represents a specifically claimed  
CC acidic peptide for use in the composition. The composition is used for  
CC transfection in vitro, in vivo or ex vivo, particularly for gene  
CC therapy. A typical application is production of tumour vaccines from  
CC autologous cells transinfected with a cytokine gene. The acidic peptide  
CC destabilises the endosomal membrane and significantly increases  
CC transfection efficiency, e.g. by as much as 1000 times, equivalent to  
CC that achieved with the optimal amount of cationic lipid. The acidic  
CC peptide also makes the transfection complex less sensitive to serum.  
XX  
XX  
SQ Sequence 23 AA:  
  
Query Match 55.1%; Score 134; DB 18; Length 23;  
Best Local Similarity 100.0%; Pred. No. 7.7e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 GGLFGAIGFTENGEGMIDGWY 24  
Db |||||  
1 glfgalagftengwegmidgwy 23  
  
RESULT 8  
AAR98016  
ID AAR98016 standard; peptide; 25 AA.  
XX  
XX  
AC AAR98016;  
XX  
XX  
DT 12-NOV-1996 (first entry)  
XX  
XX  
DE Fusogenic peptide derived from HA-2 spike glycoprotein.  
XX  
XX  
KW Fusogenic peptide; spike glycoprotein; enveloped virus; vaccine; EMDPC;  
KW endosome membrane disruption promoting component; cationic polyamine;  
KW multifunctional molecular complex; foodstuff; herbicide; insecticide;  
KW plant growth regulator; miticide; rodenticide; parasiticide;  
KW nematocide; immunisation; pathogen; therapy; autoimmune disease;  
KW hyperproliferating disease.  
XX  
XX  
OS Synthetic.  
OS  
XX  
XX  
PN WO9610038-A1.  
XX  
XX  
PD 04-APR-1996.  
XX  
XX  
PF 28-SEP-1995; 95WO-US12502.  
XX  
XX  
PR 28-SEP-1994; 94US-0314060.  
XX  
XX  
PA (APOL-) APOLLON INC.  
XX  
XX  
PI Boutin RH;  
XX  
XX  
DR WPI; 1996-200887/20.  
XX  
XX  
PT New complexes for nucleic acid transfer to target cells - comprising  
PT a nucleic acid compsn. and a cationic poly:amine with an endosome  
PT membrane disruption component.  
XX  
XX  
PS Disclosure; Page 30; 138pp; English.

XX AAR98010-R98041 represent fusogenic peptides derived from spike  
CC glycoproteins of enveloped viruses. These fusogenic peptides act as  
CC endosome membrane disruption promoting components (EMDPCs). These  
CC sequences were used in a transfer moiety which makes up part of the  
CC multifunctional molecular complex of the invention. The multifunctional  
CC molecular complex is used for the transfer of a nucleic acid composition  
CC to a target cell, and comprises the nucleic acid composition and a  
CC transfer moiety comprising one or more cationic polyamine components  
CC bound to the nucleic acid composition. There are several different  
CC cationic polyamines that can be used in the complex, one of which  
CC contains one of these sequences linked via a bridging group to the main  
CC body of the cationic polyamine. The products can be used for delivering  
CC therapeutic agents, vaccines, foodstuffs, herbicides, plant growth  
CC regulators, insecticides, miticides, rodenticides, fungicides,  
CC parasiticides or nematocides. They can also be used for immunising an  
CC individual against a pathogen or for treating an autoimmune disease or  
CC hyperproliferating disease. The complexes provide for a high level of  
CC transfection and expression of the nucleic acid molecules in target  
CC cells.

SO Sequence 25 AA:

Query Match 55.1%; Score 134; DB 17; Length 25;  
Best Local Similarity 100.0%; Pred. No. 8.5e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLEFGAIGFIENGWEGMIDGWYG 24  
Db 1 gllfgalagfiengwvgnldgwyg 23

RESULT 9

AAR26005 AAR26005 standard; peptide; 26 AA.

XX AAR26005;

XX 26-JAN-1993 (first entry)

DE Influenza fusion peptide #2.

XX Haemagglutinin; cysteine; liposome.

XX Synthetic.

XX EP497997-A.

XX 12-AUG-1992.

XX 02-FEB-1991; 91EP-0101414.

XX 02-FEB-1991; 91EP-0101414.

XX (NIKA-) NIKA HEALTH PROD LTD.

XX Gluck R, Herrmann P, Klein P;

XX WPI; 1992-270078/73.

XX Drug-conty. phospholipid bi:layer vesicle with cell-specific

XX markers on membrane - where markers have at least 90 per cent

XX biological activity, used as pharmaceuticals against e.g. cancer

XX and AIDS

XX Disclosure; Fig 2; 13pp; English.

XX The sequences given in AAR26004-19 are influenza fusion peptides,  
CC derived from the influenza virus haemagglutinin gene which were  
CC used for the preparation of synthetic membrane vesicles. The  
CC arrangement of at least one, pref. three cysteine residues at one  
CC end of these peptides has been found useful for the fusion activity,

CC for the fusion of the liposome to the target cell membrane. The  
CC liposomes produced using these peptides can contain at least one  
CC active drug and can be used to target cells infected with cancer  
CC or AIDS.

SO Sequence 26 AA:

Query Match 55.1%; Score 134; DB 13; Length 26;  
Best Local Similarity 100.0%; Pred. No. 8.8e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLEFGAIGFIENGWEGMIDGWYG 24  
Db 1 gllfgalagfiengwvgnldgwyg 23

RESULT 10

AAM34270 AAM34270 standard; peptide; 26 AA.

XX AAM34270;

XX 14-MAY-1998 (first entry)

DE Synthetic lipid vesicle fusion peptide 2.

KW Virosome; vesicle; cationic lipid bilayer; viral fusion peptide;

KW drug delivery system; membrane; gene therapy; diagnosis; treatment;

KW cancer; leukaemia; viral infection.

XX Synthetic.

XX WO9741834-A1.

XX 04-MAY-1997; 97WO-EP02268.

XX 08-MAY-1996; 96EP-0107282.

XX (NIKA-) NIKA HEALTH PROD LTD.

XX Glueck R, Klein P, Waelte ER;

XX WPI; 1997-558673/51.

XX Vesicle with cationic lipid bilayer that includes viral fusion

XX peptide - used for delivery of genetic material to cells, especially

XX for gene therapy of cancer, leukaemia and viral infections

XX Disclosure; Page 10; 52pp; English.

XX Peptides AAM34269-W34284 represent novel lipid vesicles with positively

XX charged lipid bilayer membranes composed of a cationic and/or

XX polycationic lipid and at least one natural or synthetic viral fusion

XX peptide integrated in, or covalently linked to, the membrane. Such

XX peptides are used as drug delivery systems, preferably for

XX (non)-specific delivery of genetic material to target cells or tissues,

XX particularly for diagnosis, treatment (especially antisense treatment)

XX of cancer, leukaemia and viral infections in humans or animals. Genetic

XX material is delivered, without infection, to resting or proliferating

XX cells, in vitro or in vivo. When the genetic material is an antisense  
CC molecule, it is targeted to mRNA encoding a (proto)oncogene. The  
CC continuous lipid layer does not leak. The peptides do not need to fuse  
CC with, or destabilise, plasma membranes in order to enter the cytoplasm,  
CC since the fusion peptide ensures cell penetration by endocytosis (after  
CC which fusion of the vesicle and endosomal membrane occurs). The genetic  
CC material thus has a greater chance of reaching the nucleus before it is  
CC degraded or expelled. Transfer of the material is 1000-20000 times more  
CC efficient than when using liposomes or conventional virosomes, so  
CC smaller doses can be used, avoiding possible toxicity associated with the  
CC genetic material.



XX Sequence 26 AA;  
SQ

Query Match 55.1%; Score 134; DB 18; Length 26;  
Best Local Similarity 100.0%; Pred. No. 8.8e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIEGMEGMDGWG 24  
Db 1 glfgaiaagfiengwegmidgwgy 23

RESULT 11  
AAB70091  
ID AAB70091 standard; peptide; 26 AA.

XX AAB70091;

DT 14-MAY-2001 (first entry)

DE Cationic virosome crosslinker #2 for polypeptide binding.

XX Cationic virosome; cytosolic; gene therapy; lipid bilayer vesicle;

KW viral haemagglutinin; drug delivery; cancer; leukaemia; viral infection;

XX crosslinker.

CS Unidentified.

XX NZ504444-A.

PN 24-NOV-2000.

PD 10-MAY-2000; 2000NZ-0504444.

PE 10-MAY-2000; 2000NZ-0332666.

PR (NIKA-) NIKA HEALTH PROD LTD.

PA Watti ER, Gluck R, Klein P;

PI WPI: 2001-233042/24.

PT Lipid bilayer vesicle, useful for delivering drugs to target cells such

PS as cancer, leukemic, or virally infected cells, comprises viral

XX glycoproteins in positively charged membrane

CC Disclosure; Page 14; 41pp; English.

CC The present sequence is a crosslinker which is capable of linking to the

CC surface of a novel cationic virosome and is capable of binding

CC polypeptides. The cationic virosome is a lipid bilayer vesicle comprising

CC a membrane with a net positive charge and contains 5-30 weight % based on

CC total lipids, 1,3-di-oleoyloxy-2-(6-carboxy-spermyl)-propyl-amine

CC (DOSPER) together with other lipids and at least one active fusogenic

CC peptide. The fusogenic peptide is a viral haemagglutinin that causes the

CC vesicle to be internalised by target cells through phagocytosis or

CC endocytosis. The virosome is useful for delivering a desired drug or

CC substance, preferably a nucleic acid, to target cells (resting or

CC proliferating mammalian cells) such as cancer cells, leukemic cells or

CC virally infected cells in vitro, in diagnostic or medical applications

CC and for the manufacture of medicament for carrying out non-infectious

CC gene therapy.

Db 1 glfgaiaagfiengwegmidgwgy 23

RESULT 12  
AAR98024  
ID AAR98024 standard; peptide; 28 AA.

XX AAR98024;

DT 12-NOV-1996 (first entry)

DE Fusogenic peptide derived from HA-2 X:31 spike glycoprotein.

XX Fusogenic peptide; spike glycoprotein; enveloped virus; vaccine; EMDPC;

KW endosome membrane disruption promoting component; cationic polyamine;

KW multifunctional molecular complex; foodstuff; herbicide; insecticide;

KW plant growth regulator; miticide; rodenticide; fungicide; parasiticide;

KW nematocide; immunisation; pathogen; therapy; autoimmune disease;

XX hyperproliferating disease.

OS Synthetic.

XX WO9610038-A1.

PN 04-APR-1996.

PD 28-SEP-1995; 95WO-US12502.

PE 28-SEP-1994; 94US-0314060.

PR (APOL-) APOLLON INC.

PA Boutin RH;

PI WPI: 1996-200887/20.

PT New complexes for nucleic acid transfer to target cells - comprising

PS a nucleic acid compsn. and a cationic poly:amine with an endosome

XX membrane disruption component.

CC Disclosure; Page 30; 138pp; English.

CC AAR98010-R98041 represent fusogenic peptides derived from spike

CC glycoproteins of enveloped viruses. These fusogenic peptides act as

CC endosome membrane disruption promoting components (EMDPCs). These

CC sequences were used in a transfer moiety which makes up part of the

CC multifunctional molecular complex of the invention. The multifunctional

CC molecular complex is used for the transfer of a nucleic acid composition

CC to a target cell, and comprises the nucleic acid composition and a

CC transfer moiety comprising one or more cationic polyamine components

CC bound to the nucleic acid composition. There are several different

CC cationic polyamines that can be used in the complex, one of which

CC contains one of these sequences linked via a bridging group to the main

CC body of the cationic polyamine. The products can be used for delivering

CC therapeutic agents, vaccines, foodstuffs, herbicides, plant growth

CC regulators, insecticides, miticides, rodenticides, fungicides,

CC parasiticides or nematocides. They can also be used for immunising an

CC individual against a pathogen or for treating an autoimmune disease or

CC hyperproliferating disease. The complexes provide for a high level of

Db 1 glfgaiaagfiengwegmidgwgy 23

Query Match 55.1%; Score 134; DB 17; Length 28;  
Best Local Similarity 100.0%; Pred. No. 9.5e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIEGMEGMDGWG 24  
Db 1 glfgaiaagfiengwegmidgwgy 23

RESULT 13  
AAV89680  
ID AAV89680 standard; peptide; 30 AA.  
XX  
AC AAV89680;  
XX  
DT 23-MAY-2000 (first entry)  
XX  
DE Core polypeptide fragment T No. 1242.  
XX  
KM Retrovirus; hybrid polypeptide; enhancer; gp41; envelope protein; HIV-1;  
KM HIV-2; SIV; pharmacokinetic; half-life; growth factor; cytokine; viral;  
KM anti-fusogenic; differentiation factor; interleukin; interferon;  
KW colony stimulating factor; hormone; angiogenic factor.  
XX  
OS Unidentified.  
XX  
PN WO959615-A1.  
XX  
PD 25-NOV-1999.  
XX  
PF 20-MAY-1999; 99WO-US11219.  
XX  
PR 20-MAY-1998; 98US-0082279.  
XX  
PA (TRIM-) TRIMERIS INC.  
XX  
PI Barney S, Guthrie KI, Merutka G, Anwer MK, Lambert DM;  
XX  
DR WPI; 2000-136792/12.  
XX  
PT A new hybrid polypeptide with enhanced pharmacokinetic properties  
PT comprises enhancer sequence -  
XX  
PS Disclosure; Page 42; 124pp; English.  
XX  
CC The invention relates to hybrid polypeptides comprising enhancer peptide  
CC sequence linked to core polypeptides. The enhancer polypeptides are  
CC derived from various retroviral envelope (gp41) protein sequences,  
CC especially from HIV-1, HIV-2 and SIV. The enhancer peptides enhance the  
CC pharmacokinetic properties such as increasing the half-life of any core  
CC polypeptide that they are linked to. The core polypeptides are any  
CC polypeptide that may be introduced into a living system and that can  
CC function as a pharmacologically useful peptide for the treatment or  
CC prevention of a disease. The core polypeptides are bioactive peptides  
CC selected from a growth factor, cytokine, differentiation factor,  
CC interleukin, interferon, colony stimulating factor, hormone or  
CC angiogenic factor. The peptides of the invention can be used for  
CC inhibiting viral infection and can be used in anti-viral and  
CC anti-fusogenic treatments. Sequences AAY8651-Y9005 represent core  
CC polypeptide fragments that can be used in the invention. Some sequences  
CC among those indicated also comprise enhancer fragments at terminal ends  
CC and form hybrid polypeptides.  
XX  
SQ Sequence 30 AA;  
OY  
Query Match 55.1%; Score 134; DB 21; Length 30;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2 GLFGAIAGFIENGWEGMIDGWYG 24  
DB 1 glfgaiaagiengwegmidgwyg 23

RESULT 14  
ABB01088  
ID ABB01088 standard; Peptide; 30 AA.  
XX  
AC ABB01088;  
XX

DT 03-JAN-2002 (first entry)  
XX  
DE Viral DP178/107-like region peptide T1242.  
XX  
KM Human immunodeficiency virus; HIV; respiratory syncytial virus; RSV;  
KW virucide; heptad repeat region; transmembrane protein; gp41; HR1; HR2;  
KW infection.  
XX  
OS Viridae.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 1  
FT /note= "N-terminal is substituted by Ac"  
FT Modified-site 30  
FT /note= "C-terminal amide"  
XX  
XX WO200164013-A2.  
XX  
PN 07-SEP-2001.  
XX  
PD 07-FEB-2001; 2001WO-US03988.  
XX  
PF 29-FEB-2000; 2000US-0515965.  
XX  
PR (TRIM-) TRIMERIS INC.  
XX  
PA Antczak JB, Delmedico MK, Erickson JB, Lambert DM, Sista P;  
XX  
PI WPI; 2001-514829/56.  
XX  
DR Heptad repeat region peptide analogs useful for inhibiting virus/cells  
XX  
PT fusion, useful for treating HIV and Respiratory Syncytial Virus  
PT infection -  
XX  
PS Disclosure; Page 55; 587pp; English.  
XX  
CC The invention relates to isolated analogues of the heptad repeat region  
CC peptides DP178 and DP107. DP178 and DP107 correspond to amino acids  
CC 638-673 (heptad repeat region HR2) and 558-595 (heptad repeat region  
CC HR1) respectively, of HIV-1LAI transmembrane protein gp41. The HR1  
CC and HR2 regions of proteins interact non-covalently with each other  
CC and/or with peptides derived from them. This interaction is required for  
CC normal infectivity of viruses such as RSV and HIV. The heptad  
CC repeat region peptide analogues may be used to inhibit respiratory  
CC syncytial virus (RSV) infection in a cell. They may also be used to  
CC inhibit HIV infection. The present sequence is a peptide provided in  
CC the specification.  
XX  
SQ Sequence 30 AA;  
OY  
Query Match 55.1%; Score 134; DB 22; Length 30;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2 GLFGAIAGFIENGWEGMIDGWYG 24  
DB 1 glfgaiaagiengwegmidgwyg 23

RESULT 15  
ABB02537  
ID ABB02537 standard; Peptide; 30 AA.  
XX  
AC ABB02537;  
XX  
DT 03-JAN-2002 (first entry)  
XX  
DE Viral core polypeptide, SEQ ID NO: 1064.  
XX  
KM Human immunodeficiency virus; HIV; respiratory syncytial virus; RSV;  
KW virucide; heptad repeat region; transmembrane protein; gp41; HR1; HR2;  
KW infection.

XX OS Viridiae.  
 XX PN W0200164013-A2.  
 XX PD 07-SEP-2001.  
 XX FE 07-FEB-2001; 2001MO-US03988.  
 XX PR 29-FEB-2000; 2000US-0515965.  
 XX PA (TRIM-) TRIMERIS INC.  
 XX PI Antczak JB, Delmedico MK, Erickson JB, Lambert DM, Sista P;  
 XX DR WPI; 2001-514829/56.  
 XX PT Heptad repeat region peptide analogs useful for inhibiting virus/cells  
 PT fusion, useful for treating HIV and Respiratory Syncytial Virus  
 PT infection -  
 XX PS Disclosure; Page 448; 587pp; English.  
 CC CC The invention relates to isolated analogues of the heptad repeat region  
 CC peptides DP178 and DP107. DP178 and DP107 correspond to amino acids  
 CC 638-673 (heptad repeat region HR2) and 558-595 (heptad repeat region  
 CC HR1) respectively of HIV-1/IIA1 transmembrane protein gp41. The HR1  
 CC and HR2 regions of proteins interact non-covalently with each other  
 CC and/or with peptides derived from them. This interaction is required for  
 CC normal infectivity of viruses such as RSV and HIV. The heptad  
 CC repeat region peptide analogues may be used to inhibit respiratory  
 CC syncytial virus (RSV) infection in a cell. They may also be used to  
 CC inhibit HIV infection. The present sequence is a peptide provided in  
 CC the specification.  
 XX SO Sequence 30 AA;

Query Match 55.1%; Score 134; DB 22; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 1e-09;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GTFGAIAGFIENGWEGMTDGTG 24  
 ||||||||||||||||||||  
 Db 1 gtfgaiaagiengwegmtdgtwyg 23

Search completed: July 1, 2002, 06:15:25  
 Job time: 512 sec

! FINDPATTERNS on geneseqp: \* allowing 0 mismatches

! C(R,K){20} July 1, 2002 07:07 ..

1 AA45801 ck: 9500 len: 39 ! Aa45801 One chain of a bombesin dimer. 6/1

C(R,K){20}  
C(R,K){20}  
1: CXXXXXXXXXXXXXXXXX XGGGQ

1 AAB13780 ck: 7317 len: 21 ! Aab13780 Soluble peptide antigen PK. 11/200

C(R,K){20}  
C(R,K){20}  
1: CXXXXXXXXXXXXXXXXX

1 AAU18238 ck: 5509 len: 58 ! Aau18238 Novel human DNA-binding protein #8

C(R,K){20}  
C(R,K){20}  
35: KFYFV CXXXXXXXXXXXXXXXXX KKK

1 AA003766 ck: 8808 len: 81 ! Aa003766 Human polypeptide SEQ ID NO 17658

C(R,K){20}  
C(R,K){20}  
48: LTTTA CXXXXXXXXXXXXXXXXX KKKKK

1 AA011210 ck: 863 len: 70 ! Aa011210 Human polypeptide SEQ ID NO 25102

C(R,K){20}  
C(R,K){20}  
30: IDLCL CXXXXXXXXXXXXXXXXX KKKKK

Databases searched:  
EMBL, Release 7.0, Released on 28Apr2002, Formatted on 2May2002

Total finds: 5  
Total length: 114,001,827  
Total sequences: 766,495  
CPU time: 07:03.10

Seq #6

11AA\_SEQUENCE 1.0  
ID AAW45801 standard; peptide; 39 AA.  
XX  
AC AAW45801;  
XX  
DT 25-JUN-1998 (first entry)  
XX  
DE One chain of a bombesin dimer.  
XX  
KW Alpha-melanocyte stimulating hormone; alpha-MSH; receptor agonist;  
KW alpha-MSH-AMT; bombesin; dimer; bivalent agonist; disulphide bond;  
KW G-protein coupled receptor.  
XX  
OS Synthetic.  
XX  
FH Key 1 Location/Qualifiers  
FT Cross-links 1 /note= "This residue is disulphide bonded to the  
FT corresponding Cys residue of an identical chain  
FT to form a dimer."  
FT Modified-site 28 /note= "Epsilon-amlnohexanoic acid"  
XX  
PN WO9803632-A1.  
XX  
PD 29-JAN-1998.  
XX  
PE 23-JUL-1997; 97WO-US12911.  
XX  
PR 24-JUL-1996; 96US-0686934.  
XX  
PA (UYVA ) UNIV YALE.  
XX  
PI Carriers MD, Lerner MR;  
XX  
DR WPI; 1998-120757/11.  
XX  
DT 1998-120757/11.  
XX  
PT Bivalent agonist of G-protein coupled receptors containing two  
PT ligand domains - bonded to molecular backbone, for treatment of  
PT hypotension, promotion of skin tanning etc., also for delivering  
PT drugs and gene therapy vectors to selected cells  
XX  
PS Claim 41; Page 48; 71pp; English.  
XX  
CC This sequence represents one of two identical chains disulphide bonded  
CC to form a bombesin dimer. The invention relates to bivalent agonists,  
CC with affinity for at least 1 G-protein coupled receptor (GPCR). The  
CC bivalent agonists comprise: (a) two ligand domains (LD), individually  
CC agonists or antagonists for GPCR, spaced 40-250 Angstrom apart, and (b)  
CC a molecular backbone (MB) covalently bound to LD. The bivalent agonists  
CC are useful in human or veterinary medicine as carriers for drugs or gene  
CC therapy vectors, allowing these to be endocytosed by GPCR-expressing  
CC cells. They can also be used e.g. to treat hypertension (angiotensin-  
CC based LD); to increase levels of luteinising hormone (LH), using LD  
CC derived from LH-releasing hormone, or to promote skin tanning (LD based  
CC on alpha -melanocyte-stimulating hormone, MSH). The bivalent agonists  
CC are administered orally, by injection or topically. Typical doses for  
CC skin tanning are 1-4000 (especially 30-100) mu mole/kg systemically or  
CC the bivalent agonists are administered topically in a composition  
CC containing 0.001-10 (especially 1) mM. Where both LD are agonists, the  
CC bivalent agonist has a synergistically higher activity than two  
CC individual agonist ligands, and where at least 1 is an antagonist the  
CC effect is stimulatory. The bivalent agonists are active at lower  
CC concentrations than known agonists so should avoid toxicity problems.  
XX  
SQ Sequence 39 AA;  
AAW45801 Length: 39 July 1, 2002 14:50 Type: P Check: 9500 ..  
1 CKRKKKKKK KKKKKKKKK KXGGGQRRG NQWAVGHM  
11AA\_SEQUENCE 1.0  
ID AAB13780 standard; peptide; 21 AA.

XX  
AC AAB13780;  
XX  
DT 10-NOV-2000 (first entry)  
XX  
DE Soluble peptide antigen pK.  
XX  
KW pK peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;  
KW major histocompatibility complex class 1; MHC class 1; antigen; tumour;  
KW prostate; breast; multiple myeloma.  
XX  
OS Unidentified.  
XX  
PN WO200035949-A1.  
XX  
PD 22-JUN-2000.  
XX  
PE 14-DEC-1999; 99WO-US29724.  
XX  
PR 14-DEC-1999; 98US-0112324.  
XX  
PA (DEND-) DENDREON CORP.  
XX  
PI Laus R, Hakim I, Vidovic D;  
XX  
DR WPI; 2000-442365/38.  
XX  
DT 2000-442365/38.  
XX  
PT Antigens modified by the covalent addition of a peptide that  
PT facilitates entry into antigen presenting cells, useful for producing  
PT compositions for immunising against tumors and pathogens -  
XX  
PS Claim 2; Page 26; 34pp; English.  
XX  
CC The present invention relates to compositions of modified soluble protein  
CC antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)  
CC response i.e. a major histocompatibility complex (MHC) class I molecule  
CC response. The protein antigen is modified by the covalent addition of a  
CC peptide sequence which facilitate entry of the antigen into antigen  
CC presenting cells (APCs). The present sequence is one such peptide  
CC sequence which can be used to modify the soluble antigens. The present  
CC sequence is peptide pK. The modified antigen composition may be used for  
CC immunising against, or treating a tumour e.g. prostate and breast  
CC carcinoma or multiple myeloma, or pathogen in mammals.  
XX  
SQ Sequence 21 AA;  
AAB13780 Length: 21 July 1, 2002 14:50 Type: P Check: 7317 ..  
1 CKRKKKKKK KKKKKKKKK K  
11AA\_SEQUENCE 1.0  
ID AAU18238 standard; Protein; 58 AA.  
XX  
AC AAU18238;  
XX  
DT 21-NOV-2001 (first entry)  
XX  
DE Novel human DNA-binding protein #85.  
XX  
KW Human; DNA-binding protein; histone; chromo domain protein;  
KW chromatin organisation modifier; Y-box binding protein;  
KW DNA organisation; gene transcription; malignant disease;  
KW autoimmune disorder; rheumatic disease; genetic abnormality;  
KW infectious disease; neurological disorder; gene therapy;  
KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;  
KW cytostatic.  
XX  
OS Homo sapiens.  
XX  
PN WO200155162-A1.  
XX  
PD 02-AUG-2001.

```
PF 17-JAN-2001; 2001WO-0501305.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0216135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 23-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
XX

PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 13-OCT-2000; 2000US-0239939.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249247.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250330.
PR 05-DEC-2000; 2000US-0251093.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251997.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA
XX PI Rosen CA, Barash SC, Ruben SM;
XX MPI: 2001-465557/50.
XX DR N-PSDB; AAS29114.
XX
```

PT Nucleic acid molecules encoding human secreted chromosomal binding  
PT proteins, used in preventing, treating or ameliorating a disorder, e.g.  
PT Alzheimer's and Parkinson's diseases and cancers -  
PS  
PS Claim 11: SEQ ID NO 223; 561pp; English.  
XX  
CC The present invention relates to the isolation of novel DNA-binding  
CC proteins, and cDNA (AA829030-AA829157) and genomic sequences encoding  
CC for these proteins. DNA-binding proteins such as histones, chromo  
CC (chromatin organisation modifier) domain proteins, and Y-box binding  
CC proteins may contribute to diseases resulting from aberrant DNA  
CC organisation and/or gene transcription. The sequences of the invention  
CC are useful in screening assays to identify antagonists and/or agonists  
CC that may enhance or block activities mediated by DNA-binding proteins.  
CC Blockers of DNA-binding proteins may be useful in treating disorders  
CC such as malignant diseases (e.g. cancer), autoimmune disorders  
CC (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid  
CC arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious  
CC diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's  
CC disease). The polynucleotide sequences of the invention may also be  
CC used in gene therapy. AAU18154-AAU18281 represent novel DNA-binding  
CC proteins.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
CC  
SQ Sequence 58 AA:  
AAU18238 Length: 58 July 1, 2002 14:50 Type: P Check: 5509 ..  
1 TYLECEHNSL VNSKCLTVVL SRCISVCLNK FYVCKKKKK KKKKKKKKK  
51 KKKKKKKK  
11AA\_SEQUENCE 1.0  
ID AA003766 standard; Protein: 81 AA.  
XX  
AC AA003766;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 17658.  
XX  
KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KM tissue growth factor; immunomodulatory; cancer; leukaemia;  
KM nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PE 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
XX  
DR WPI: 2001-514838/56.  
DR N-PSDB; AAI83697.  
XX  
PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
PS Claim 20: SEQ ID NO 17658; 1399pp + Sequence Listing; English.  
XX  
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and

CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation and which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activity/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
CC  
SQ Sequence 81 AA:  
AA003766 Length: 81 July 1, 2002 14:50 Type: P Check: 8808 ..  
1 GLNQTQLRKI LAYSSITHG XIIVLPYNP NITILNLTIV ILLTTACKK  
51 KKKKKKKKK KKKKKKKKK KKKKKGGG A  
11AA\_SEQUENCE 1.0  
ID AA011210 standard; Protein: 70 AA.  
XX  
AC AA011210;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 25102.  
XX  
KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KM tissue growth factor; immunomodulatory; cancer; leukaemia;  
KM nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PE 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
XX  
DR WPI: 2001-514838/56.  
DR N-PSDB; AAI91141.  
XX  
PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
PS Claim 20: SEQ ID NO 25102; 1399pp + Sequence Listing; English.  
XX  
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation and which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activity/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 70 AA;

AA01210 Length: 70 July 1, 2002 14:50 Type: P Check: 863 ..

1 YTHRTVFM CMNGLKDNV DKXTIDLCLC KKKKKKKKK KKKKKKKKK

51 KKKKKKKKK KKKPOGGGA

---



```

!!SEQUENCE LIST 1.0
! FINDPATTERNS on geneseqp:* allowing 0 mismatches
!      1 C(R,K){20}
July 1, 2002 14:38 ..

GENESEQP1906:AAW45801 ck: 9500 len: 39 finds: 1 ! Aaw45801 One chain of a bomb
GENESEQP2000:AAI13780 ck: 7317 len: 21 finds: 1 ! Aai13780 Soluble peptide anti
GENESEQP2001:AAU18238 ck: 5509 len: 58 finds: 1 ! Aau18238 Novel human DNA-bind
GENESEQP2001:AAO03766 ck: 8808 len: 81 finds: 1 ! Aao03766 Human polypeptide SE
GENESEQP2001:AAO11210 ck: 863 len: 70 finds: 1 ! Aao11210 Human polypeptide SE

\\End of list

Databases searched:
EMBL, Release 7.0, Released on 28Apr2002, Formatted on 2May2002

Total finds: 5
Total length: 114,001,827
Total sequences: 766,495
CPU time: 09:53.67

```

! FINDPATTERNS on pir:\* allowing 0 mismatches

! 1 C(R,K){20}

July 1, 2002 06:14 ..

Databases searched:

NRRL Release 71.0, Released on 15Feb2002, Formatted on 19Feb2002

Total finds:

0

Total length:

96,089,334

Total sequences:

283,138

CPU time:

04:30.92

! FINDPATTERNS ON SWP:\* allowing 0 mismatches

! 1 C(R,K)(20)

July 1, 2002 06:15 ..

Databases searched:

SWISS-PROT, Release 40.1, Released on 11Feb2002, Formatted on 19Feb2002  
SPRMBL, Release 19.0, Released on 14Dec2001, Formatted on 19Feb2002

Total finds:

0

Total length:

211,714,479

Total sequences:

667,446

CPU time:

10:17.12

! FINDPATTERNS on pir: \* allowing 0 mismatches

1 (R,K){20}

July 1, 2002 06:48 ..

T49173 ck: 4143 len: 517 1 hypothetical protein T20N10.250 - Arabidops

1 (R,K){20}

444: FERVG KKKKKKKKKKKKKKKKKKK KKKIR

(K){20}

445: ERYGK KKKKKKKKKKKKKKKKKKK KKKIRL

(K){20}

446: RVGKK KKKKKKKKKKKKKKKKKKK KIRLN

(K){20}

447: VGKKK KKKKKKKKKKKKKKKKKKK IRLNF

S58321 ck: 1384 len: 126 1 Probable membrane protein YOR309c - yeast

(R,K){20}

53: RKRRT RKRRRRRRRRRRRRRRRRR KRSPR

(R,K){20}

54: KRRTT KRRRRRRRRRRRRRRRRRR RSPRR

(R,K){20}

55: RRTTR RRRRRRRRRRRRRRRRRR SPRRR

T46395 ck: 7330 len: 380 1 hypothetical protein DKFZp434I1120.1 - huma

(R,K){20}

355: NLLQ KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

356: LLLQ KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

357: LLOK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

358: LOKK KKKKKKKKKKKKKKKKKKK KKK

(K){20}

359: QKKK KKKKKKKKKKKKKKKKKKK KK

(K){20}

360: KKKK KKKKKKKKKKKKKKKKKKK K

(K){20}

361: KKKK KKKKKKKKKKKKKKKKKKK

I52523 ck: 8048 len: 215 1 nucleoporin p62 homolog - rat (fragment)

(R,K){20}

35: CEPLE KKKKKKKKKKKKKKKKKKK KKTGD

(K){20}

36: EFLEK KKKKKKKKKKKKKKKKKKK KTGDN

(K){20}

37: FLEK KKKKKKKKKKKKKKKKKKK TGDNA

NBRF, Release 71.0, Released on 15Feb2002, Formatted on 19Feb2002  
Total finds: 17  
Total length: 96,089,334  
Total sequences: 283,138  
CPU time: 09:02.03

Databases searched:

!!SEQUENCE\_LIST 1.0  
! FINDPATTERNS on pir:\* allowing 0 mismatches

1 (R,K)(20)

July 1, 2002 14:57 ..

PIR2:T49173 CK: 4143 len: 517 finds: 4 1 hypothetical protein T20N10.25  
PIR2:S58321 CK: 1384 len: 126 finds: 3 1 probable membrane protein YOR3  
PIR2:T46395 CK: 7330 len: 380 finds: 7 1 hypothetical protein DKFZp434I  
PIR2:I52523 CK: 8048 len: 215 finds: 3 1 nucleoporin p62 homolog - rat

\\End of list

Databases searched:

NBRF, Release 71.0, Released on 15Feb2002, Formatted on 19Feb2002

Total finds: 17  
Total length: 96,089,334  
Total sequences: 283,138  
CPU time: 08:41.42

!!AA\_SEQUENCE 1.0  
 PI:T49173 - hypothetical protein T20N10.250 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 08-Dec-2000  
 C:Accession: T49173  
 R:D'Angelico, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.;  
 Rudd, S.; Lemke, K.; Mayer, R.F.X.; Queirer, F.; Salanoubat, M.  
 Submitted to the Protein Sequence Database, April 2000  
 A:Reference number: 225017  
 A:Accession: T49173  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-517 <DAN>  
 A:Cross-references: EMBL:AL353032; GSPDB:GN00061; ATSP:T20N10.250  
 A:Experimental source: cultivar Columbia; BAC clone T20N10  
 C:Genetics:  
 A:Gene: ATSP:T20N10.250  
 A:Map position: 3  
 A:Introns: 312/3; 359/3; 444/3  
 C:Superfamily: Arabidopsis thaliana hypothetical protein F17J16.30

T49173 Length: 517 July 1, 2002 15:06 Type: P Check: 4143 ..

1 MDLFSSLPNE LLYHILSFLS TKKALATSVL SKRMNLEAF VPYLEFDDSV  
 51 FLHPEERKAE KEGILQSFMD FVDRVLDLHG DSLIKTFSLK CKTGVSDDHV  
 101 DMRICVVLAR GVSDDLDFID FRDLVSLPHE VGVSRITLVYL RVGSESDLYW  
 151 WQKFLCPML KTVLDSCWL CIGQFOILL ACPALBELDM TMTWKDSNV  
 201 TVSSSLIKEL TIDHGCCSV VNKSLSFDA PSLYFYCYD SLADYDPQVN  
 251 LKNIVEAQIN LLTQAOQIEQ VRALNEMLV ADDVEPGLN AWKLITGLRN  
 301 VQQLVISPOT LEVLSRCEG MPVFNKLVY SIMSDMNRW QAMVLLRNC  
 351 PLELTILIEG LLHYATDKCG DVDCDISRDY KDHSLTSCV KKLQIYEFEG  
 401 TIRELEMIKH FLKIFPCLKE MDIYAHENSH TLEKDPITFE RVGKRRKKKK  
 451 KKKKKKKKK KKKKKIKLN FKPVNKTGEF LKRLADKICF IPQCLEFLDV  
 501 DSSLGELALL AMDSRPS

!!AA\_SEQUENCE 1.0  
 PI:S58321 - probable membrane protein YOR309C - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: hypothetical protein 06105  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 21-Jul-2000  
 C:Accession: S58321; S67215; S71989  
 R:Pearson, B.M.; Hernandez, Y.; Wolf, S.S.; Kalogeropoulos, A.; Schweizer, M.  
 Submitted to the EMBL Data Library, August 1995  
 A:Reference number: S58318  
 A:Accession: S58321  
 A:Molecule type: DNA  
 A:Residues: 1-126 <PEA>  
 A:Cross-references: EMBL:X90565; NID:9940836; PID:9940840  
 R:Pearson, B.M.; Hernandez, Y.; Kalogeropoulos, A.; Schweizer, M.  
 Submitted to the Protein Sequence Database, July 1996  
 A:Reference number: S67213  
 A:Accession: S67215  
 A:Molecule type: DNA  
 A:Residues: 1-126 <PEW>  
 A:Cross-references: EMBL:Z75217; NID:91420680; PID:9252431; PID:91420681;  
 MIPS:YOR309C  
 A:Experimental source: strain S288C  
 R:Pearson, B.M.; Hernandez, Y.; Payne, J.; Wolf, S.S.; Kalogeropoulos, A.;  
 Schweizer, M.  
 Yeast 12, 1021-1031, 1996  
 A:Title: Sequencing of a 35.71 kb DNA segment on the right arm of yeast  
 chromosome XV reveals regions of similarity to chromosomes I and XIII.  
 A:Reference number: S71986; MUID:97051589

A:Accession: S71989  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-126 <PEF>  
 A:Cross-references: EMBL:X90565; NID:9940836; PID:CAA62164.1; PID:9940840  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August  
 1995  
 C:Genetics:  
 A:Map position: 15R  
 C:Keywords: transmembrane protein  
 F:3-19/Domain: transmembrane #status predicted <TM1>  
 F:107-123/Domain: transmembrane #status predicted <TM2>

S58321 Length: 126 July 1, 2002 15:06 Type: P Check: 1984 ..

1 MQLIPQRL LILNPLMMK RRRKKRRKR RERTMMKIP RILKILRRR  
 51 RTRRRRRKR RRRRRRRKR RRRSPRRKR RRRNKDAFYI LIISPSRSL  
 101 LFGFRKFSII IQCLTFSEH ILFHNL

!!AA\_SEQUENCE 1.0  
 PI:T46395 - hypothetical protein DKFZp434I1120.1 - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
 C:Accession: T46395  
 R:Ottewill, B.; Obermayer, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
 Submitted to the Protein Sequence Database, January 2000  
 A:Reference number: 223031  
 A:Accession: T46395  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-380 <AAA>  
 A:Cross-references: EMBL:AL137556  
 A:Experimental source: adult testis; clone DKFZp434I1120  
 C:Genetics:  
 A:Note: DKFZp434I1120.1

T46395 Length: 380 July 1, 2002 15:06 Type: P Check: 7330 ..

1 MGSTDSKLNK RKAIVQLTKR TOPVEATDDA FWDQFMDTA TSVQDFVALY  
 51 PAATIRAVRE EPPSNLALIC YKAVEKLVOG AESGCHSEKE KQVLNCSRL  
 101 LTRVLPYTFE DDPWKGFEWS TVPGAGRGCG EEDDEHARPL AELLALAIAD  
 151 LLFCDPFTVQ SHRRSTVDSA EDVHSLDSCG YIWEAGVGA HSPQPNYIHD  
 201 MNRRELKLL LQCFSEAMVL PPAPSGSN PMWQFCSTE NRHALPLFTS  
 251 LLNTVCAYDP VGYGIPYNEH LFSYREPLV EEAQVILVT LHDSSASSAS  
 301 PTVDGTTGT AMDADPPG ENLFVNYLSR IHREDFOFI LKGIARLNS  
 351 LLQKRRKKK KKKKKKKKK KKKKKKKKK

!!AA\_SEQUENCE 1.0  
 FI:15253 - nucleoporin p62 homolog - rat (fragment)  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 27-Feb-1997 #sequence\_revision 27-Feb-1997 #text\_change 05-Nov-1999  
 C:Accession: 15253  
 R:Mang, Z.O.; Akmal, K.M.; Kim, K.H.  
 Biol. Reprod. 51, 1022-1030, 1994  
 A:Title: An unusual nucleoporin-related messenger ribonucleic acid is present  
 in the germ cells of rat testis.  
 A:Reference number: 152523  
 A:Accession: 152523  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-215 <RES>  
 A:Cross-references: GB:S75997; NID:9913245; PIDN:AA33384.1; PID:9913246  
 A:Experimental source: testis

I52523 Length: 215 July 1, 2002 15:06 Type: P Check: 8048 ..  
1 SGGRAITSCD EDCISSLPE SLGVPKQDC EFLEKRRKK KKKRRKKKK  
51 KKKKKTGDN AKSVRQYSL KTKLEHEAE QAKVELDFIL SQKLEEDLL  
101 SPLEESVKEG SGTIYLQHAD EEREKTYKLA ENIDAQLKRM AODLKDIIEH  
151 LNMAGGPADT SDPLQOICKI LNAHMSLQW VDQSSALLOR RVEEASRYCE  
201 SRRKEQERSL RIAPD

! FINDPATTERNS on swp:\* allowing 0 mismatches

! 1 (R,K){20}

July 1, 2002 06:38 ..

Q12444 ck: 1384 len: 126 ! Q12444 saccharomyces cerevisiae (baker's ye

1 (R,K){20}

53: KRRRT RRRRRRRRRRRRRRRRRR KRSPR

54: KRRTT RRRRRRRRRRRRRRRRRR RSPRK

55: RRTTR RRRRRRRRRRRRRRRRRR SPRRR

Q9P529 ck: 291 len: 128 ! Q9P529 neurospora crassa. hypothetical 15.2

1 (R,K){20}

71: KRRNG KRRRRRRRRRRRRRRRRR KRRKK

72: RKNOK KRRRRRRRRRRRRRRRRR KRRKK

73: KNOKK KRRRRRRRRRRRRRRRRR KRRKK

74: NOKKK KRRRRRRRRRRRRRRRRR KRRKK

75: OKKKK KRRRRRRRRRRRRRRRRR KRRKK

76: KKKKK KRRRRRRRRRRRRRRRRR KRRKK

77: KKKKK KRRRRRRRRRRRRRRRRR KRRKK

78: KKKKK KRRRRRRRRRRRRRRRRR KRRKK

79: KKKKK KRRRRRRRRRRRRRRRRR KRRKE

80: KKKKK KRRRRRRRRRRRRRRRRR KRRKO

81: KKKKK KRRRRRRRRRRRRRRRRR KRRKO

82: KKKKK KRRRRRRRRRRRRRRRRR KRRKS

83: KKKKK KRRRRRRRRRRRRRRRRR KRRSR

Q9HC48 ck: 7602 len: 667 ! Q9HC48 homo sapiens (human). ctcl tumor ant

1 (R,K){20}

648: GDXTD RRRRRRRRRRRRRRRRRR

Q9H6Q7 ck: 3351 len: 720 ! Q9H6Q7 homo sapiens (human). cdna: flj21979

1 (R,K){20}

692: IVSIS KRRRRRRRRRRRRRRRRR KRRKK

693: VSISK KRRRRRRRRRRRRRRRRR KRRKK

694: SISKK KRRRRRRRRRRRRRRRRR KRRKK

695: ISKKK KRRRRRRRRRRRRRRRRR KRRKK

696: SKKKK KRRRRRRRRRRRRRRRRR KRRKK

697: KKKKK KRRRRRRRRRRRRRRRRR KRRKK

698: KKKKK KRRRRRRRRRRRRRRRRR KRRKK

699: KKKKK KRRRRRRRRRRRRRRRRR KRRKK

700: KKKKK KRRRRRRRRRRRRRRRRR KRRKK

701: KKKKK KRRRRRRRRRRRRRRRRR KRRKK

Q9H5V6 ck: 379 len: 168 ! Q9H5V6 homo sapiens (human). cdna: flj22

1

(R,K){20}

140: VREWE KRRRRRRRRRRRRRRRRR KRRKK

141: REWER KRRRRRRRRRRRRRRRRR KRRKK

142: EWEKK KRRRRRRRRRRRRRRRRR KRRKK

143: WEKKK KRRRRRRRRRRRRRRRRR KRRKK

144: EKKKK KRRRRRRRRRRRRRRRRR KRRKK

145: KKKKK KRRRRRRRRRRRRRRRRR KRRKK

146: KKKKK KRRRRRRRRRRRRRRRRR KRRKK

147: KKKKK KRRRRRRRRRRRRRRRRR KRRKK

148: KKKKK KRRRRRRRRRRRRRRRRR KRRKK

149: KKKKK KRRRRRRRRRRRRRRRRR KRRKK

Q9NT34 ck: 7330 len: 380 ! Q9NT34 homo sapiens (human). hypothetical

1

(R,K){20}

355: NLLLO KRRRRRRRRRRRRRRRRR KRRKK

356: LLLLO KRRRRRRRRRRRRRRRRR KRRKK



```

357: L0QK KKKKKKKKKKKKKKKKKKK KKKK
      {x} [20]
358: L0QK KKKKKKKKKKKKKKKKKKK KK
      {x} [20]
359: QKQK KKKKKKKKKKKKKKKKK KK
      {x} [20]
360: KKKK {x} [20]
      KKKKKKKKKKKKKKKKK K
361: KKKK {x} [20]
      KKKKKKKKKKKKKKKKK
```

```

Q951V6  cr: 7515  len: 531  ! Q951v6 macaca fascicularis (crab eating mac
      (R,K){20}
502: YKONS  KKKKKKKKKKKKKKKKKKKKK  KKKKK
      (K){20}
503: KGNSK  KKKKKKKKKKKKKKKKKKKKK  KKKKK
      (K){20}
504: GNSKK  KKKKKKKKKKKKKKKKKKKKK  KKKKK
      (K){20}
505: NSKKK  KKKKKKKKKKKKKKKKKKKKK  KKKKK
      (K){20}
506: SKKKK  KKKKKKKKKKKKKKKKKKKKK  KKKKK
      (K){20}
507: KKKKK  KKKKKKKKKKKKKKKKKKKKK  KKKKK
      (K){20}
508: KKKKK  KKKKKKKKKKKKKKKKKKKKK  KKKKK
      (K){20}
509: KKKKK  KKKKKKKKKKKKKKKKKKKKK  KKK
      (K){20}
510: KKKKK  KKKKKKKKKKKKKKKKKKKKK  KK
      (K){20}
511: KKKKK  KKKKKKKKKKKKKKKKKKKKK  K
      (K){20}
512: KKKKK  KKKKKKKKKKKKKKKKKKKKK
      (K){20}

```

```

Q9LXR2  CK: 4143  len: 517  | Q9LXR2 arabidopsis thaliana (mouse-ear cre
      (R,K){20}
      (K){20}
444: EERVG KKKKKKKKKKKKKKKKK KKKIR
      (K){20}
445: ERYGK KKKKKKKKKKKKKKKKK KKRIL
      (K){20}
446: RVGKK KKKKKKKKKKKKKKKKK KIRLN
      (K){20}
447: VGKKK KKKKKKKKKKKKKKKKKK IRLNF

```

```
Q9UG29  ck: 6094  len: 260  ! Q9UG29 arabidopsis thaliana (mouse-ear cress)
(R,K){20}
(R,K){20}
6: MDRCI RRRRRRRRRRRRRRRRRRRR RRRRR
```

7:	D0C1R	XXXXXXXXXXXXXXXXXXXX	XXXXX
8:	RC1RK	XXXXXXXXXXXXXXXXXXXX	XXXXX
9:	C1RK	XXXXXXXXXXXXXXXXXXXX	XXXXX
10:	IRKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
11:	KKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
12:	KKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
13:	KKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
14:	KKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
15:	KKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
16:	KKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
17:	KKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
18:	KKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
19:	KKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
20:	KKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
21:	KKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
22:	KKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
23:	KKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
24:	KKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
25:	KKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
26:	KKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
27:	KKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
28:	KKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
29:	KKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
30:	KKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX

31: KKKK (K){20} KKKK  
32: KKKK (K){20} KKKK  
33: KKKK (K){20} KKKK  
34: KKKK (K){20} KKKK  
35: KKKK (K){20} KKKK  
36: KKKK (K){20} KKKK  
37: KKKK (K){20} KKKK  
38: KKKK (K){20} KKKK  
39: KKKK (K){20} KKKK  
40: KKKK (K){20} KKKK  
41: KKKK (K){20} KKKK  
42: KKKK (K){20} KKKK  
43: KKKK (K){20} KKKK  
44: KKKK (K){20} KKKK  
45: KKKK (K){20} KKKK  
46: KKKK (K){20} KKKK  
47: KKKK (K){20} KKKK  
48: KKKK (K){20} KKKK  
49: KKKK (K){20} KKKK  
50: KKKK (K){20} KKKK  
51: KKKK (K){20} KKKK  
52: KKKK (K){20} KKKK  
53: KKKK (K){20} KKKK  
54: KKKK (K){20} KKKK

55: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
56: KKKK (K){20} KKKK  
57: KKKK (K){20} KKKK  
58: KKKK (K){20} KKKK  
59: KKKK (K){20} KKKK  
60: KKKK (K){20} KKKK  
61: KKKK (K){20} KKKK  
62: KKKK (K){20} KKKK  
63: KKKK (K){20} KKKK  
64: KKKK (K){20} KKKK  
65: KKKK (K){20} KKKK  
66: KKKK (K){20} KKKK  
67: KKKK (K){20} KKKK  
68: KKKK (K){20} KKKK  
69: KKKK (K){20} KKKK  
70: KKKK (K){20} KKKK  
71: KKKK (K){20} KKKK  
72: KKKK (K){20} KKKK  
73: KKKK (K){20} KKKK  
74: KKKK (K){20} KKKK  
75: KKKK (K){20} KKKK  
76: KKKK (K){20} KKKK  
77: KKKK (K){20} KKKK  
78: KKKK (K){20} KKKK  
79: KKKK (K){20} KKKK

80: KKKK (K){20} KKKK  
 81: KKKK (K){20} KKKK  
 82: KKKK (K){20} KKKK  
 83: KKKK (K){20} KKKK  
 84: KKKK (K){20} KKKK  
 85: KKKK (K){20} KKKK  
 86: KKKK (K){20} KKKK  
 87: KKKK (K){20} KKKK  
 88: KKKK (K){20} KKKK  
 89: KKKK (K){20} KKKK  
 90: KKKK (K){20} KKKK  
 91: KKKK (K){20} KKKK  
 92: KKKK (K){20} KKKK  
 93: KKKK (K){20} KKKK  
 94: KKKK (K){20} KKKK  
 95: KKKK (K){20} KKKK  
 96: KKKK (K){20} KKKK  
 97: KKKK (K){20} KKKK  
 98: KKKK (K){20} KKKK  
 99: KKKK (K){20} KKKK  
 100: KKKK (K){20} KKKK  
 101: KKKK (K){20} KKKK  
 102: KKKK (K){20} KKKK  
 103: KKKK (K){20} KKKK

104: KKKK (K){20} KKKK  
 105: KKKK (K){20} KKKK  
 106: KKKK (K){20} KKKK  
 107: KKKK (K){20} KKKK  
 108: KKKK (K){20} KKKK  
 109: KKKK (K){20} KKKK  
 110: KKKK (K){20} KKKK  
 111: KKKK (K){20} KKKK  
 112: KKKK (K){20} KKKK  
 113: KKKK (K){20} KKKK  
 114: KKKK (K){20} KKKK  
 115: KKKK (K){20} KKKK  
 116: KKKK (K){20} KKKK  
 117: KKKK (K){20} KKKK  
 118: KKKK (K){20} KKKK  
 119: KKKK (K){20} KKKK  
 120: KKKK (K){20} KKKK  
 121: KKKK (K){20} KKKK  
 122: KKKK (K){20} KKKK  
 123: KKKK (K){20} KKKK  
 124: KKKK (K){20} KKKK  
 125: KKKK (K){20} KKKK  
 126: KKKK (K){20} KKKK  
 127: KKKK (K){20} KKKK

```

128: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
129: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
130: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
131: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
132: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
133: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
134: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
135: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
136: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
137: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
138: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
139: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
140: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
141: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
142: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
143: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
144: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
145: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
146: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
147: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
148: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
149: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
150: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
151: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
152: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}

```

```

153: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
154: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
155: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
156: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
157: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
158: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
159: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
160: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
161: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
162: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
163: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
164: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
165: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
166: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
167: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
168: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
169: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
170: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
171: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
172: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
173: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
174: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
175: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
176: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}

```

177: kkkk (K){20} kkkk  
 178: kkkk (K){20} kkkk  
 179: kkkk (K){20} kkkk  
 180: kkkk (K){20} kkkk  
 181: kkkk (K){20} kkkk  
 182: kkkk (K){20} kkkk  
 183: kkkk (K){20} kkkk  
 184: kkkk (K){20} kkkk  
 185: kkkk (K){20} kkkk  
 186: kkkk (K){20} kkkk  
 187: kkkk (K){20} kkkk  
 188: kkkk (K){20} kkkk  
 189: kkkk (K){20} kkkk  
 190: kkkk (K){20} kkkk  
 191: kkkk (K){20} kkkk  
 192: kkkk (K){20} kkkk  
 193: kkkk (K){20} kkkk  
 194: kkkk (K){20} kkkk  
 195: kkkk (K){20} kkkk  
 196: kkkk (K){20} kkkk  
 197: kkkk (K){20} kkkk  
 198: kkkk (K){20} kkkk  
 199: kkkk (K){20} kkkk  
 200: kkkk (K){20} kkkk

201: kkkk kkkkkkkkkkkkkkkkk kkkk  
 202: kkkk (K){20} kkkk  
 203: kkkk (K){20} kkkk  
 204: kkkk (K){20} kkkk  
 205: kkkk (K){20} kkkk  
 206: kkkk (K){20} kkkk  
 207: kkkk (K){20} kkkk  
 208: kkkk (K){20} kkkk  
 209: kkkk (K){20} kkkk  
 210: kkkk (K){20} kkkk  
 211: kkkk (K){20} kkkk  
 212: kkkk (K){20} kkkk  
 213: kkkk (K){20} kkkk  
 214: kkkk (K){20} kkkk  
 215: kkkk (K){20} kkkk  
 216: kkkk (K){20} kkkk  
 217: kkkk (K){20} kkkk  
 218: kkkk (K){20} kkkk  
 219: kkkk (K){20} kkkk  
 220: kkkk (K){20} kkkk  
 221: kkkk (K){20} kkkk  
 222: kkkk (K){20} kkkk  
 223: kkkk (K){20} kkkk  
 224: kkkk (K){20} kkkk  
 225: kkkk (K){20} kkkk

226: KKKK (K)(20)  
KKKKKKKKKKKKKKKK KNRHH

227: KKKK (K)(20)  
KKKKKKKKKKKKKKKK NRHHI

035807 ck: 7510 len: 129 1 035807 ratus norvegicus (rat). microvascul

1 (R,K)(20)

85: VLAS (K)(20)  
KKKKKKKKKKKKKKKK KIKWE

86: LLASR (K)(20)  
KKKKKKKKKKKKKKKK IKWEG

064075 ck: 8048 len: 215 1 064075 ratus sp. nucleoporin p62 homolog P

1 (R,K)(20)

35: CEPLK (K)(20)  
KKKKKKKKKKKKKKKK KKTGD

36: EPLEK (K)(20)  
KKKKKKKKKKKKKKKK KTGDN

37: FLEKK (K)(20)  
KKKKKKKKKKKKKKKK TGDNA

Databases searched:

SWISS-PROT, Release 40.1, Released on 11Feb2002, Formatted on 19Feb2002  
SPTREMBL, Release 19.0, Released on 14Dec2001, Formatted on 19Feb2002

Total finds: 286

Total length: 211,714,479

Total sequences: 667,446

CPU time: 18:32.89

```
!!SEQUENCE LIST 1.0
! FINDPATTERNS on swp:* allowing 0 mismatches
! 1 (R,K){20}
```

July 1, 2002 15:28 ..

SP_FUN:Q12444	ck: 1384	len: 126	finds: 3	! Q12444	saccharomyces cerevisia
SP_FUN:Q9P529	ck: 291	len: 128	finds: 13	! Q9P529	neurospora crassa. hypc
SP_HUM:Q9HC48	ck: 7602	len: 667	finds: 1	! Q9HC48	homo sapiens (human). c
SP_HUM:Q9H6Q7	ck: 3351	len: 720	finds: 10	! Q9H6Q7	homo sapiens (human). c
SP_HUM:Q9H5V6	ck: 379	len: 168	finds: 10	! Q9H5V6	homo sapiens (human). c
SP_HUM:Q9NT34	ck: 7330	len: 380	finds: 7	! Q9NT34	homo sapiens (human). h
SP_OM:Q95LV6	ck: 7515	len: 531	finds: 11	! Q95LV6	macaca fascicularis (cr
SP_PL:Q9LXR2	ck: 4143	len: 517	finds: 4	! Q9LXR2	arabidopsis thaliana (t
SP_PL:Q9LQZ9	ck: 6094	len: 260	finds: 222	! Q9LQZ9	arabidopsis thaliana (m
SP_RO:Q35807	ck: 7510	len: 129	finds: 2	! Q35807	rattus norvegicus (rat)
SP_RO:Q64075	ck: 8048	len: 215	finds: 3	! Q64075	rattus sp. nucleoporin

\\End of list

Databases searched:

SWISS-PROT, Release 40.1, Released on 11Feb2002, Formatted on 19Feb2002  
 SPTREMBL, Release 19.0, Released on 14Dec2001, Formatted on 19Feb2002

Total finds: 286  
 Total length: 211,714,479  
 Total sequences: 667,446  
 CPU time: 21:25.76

```

11AA_SEQUENCE 1.0
ID Q12444 PRELIMINARY; PRT; 126 AA.
AC Q12444;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE ORF YOR309C.
GN YOR309C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycs.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Pearson B.M., Hernandez Y., Kalogeropoulos A., Schweizer M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MIRS;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=FY1679;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X90565; CAAG629.1; -.
DR EMBL; X90565; CAAG629.1; -.
DR SGD; S0005836; YOR309C.
SQ
SEQUENCE 126 AA; 16294 MW; 46EF1F4C664802C8 CRC64;

Q12444 Length: 126 July 1, 2002 15:52 Type: P Check: 1384 ..

1 MOWLIPQRL LILNPLIMK RKRKRKRKR RERETMKIP RILKLRKR
51 RRRRRRRR KRRRRKRKR RKRSPKRKR KRKKDAFYI LIIDPSRL
101 LFGFRKPSIT IQCLTFSPH ILFHNH

11AA_SEQUENCE 1.0
ID Q9P529 PRELIMINARY; PRT; 128 AA.
AC Q9P529;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 15.2 KDA PROTEIN.
GN B24H17.160.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Algu V., Hohlseil J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL356815; CAB92636.2; -.
KW Hypothetical protein.
SQ
SEQUENCE 128 AA; 15157 MW; 8C7C65C3DFB70765 CRC64;

Q9P529 Length: 128 July 1, 2002 15:52 Type: P Check: 291 ..

1 MAIISGLHH KNINRAFGH SVYSKSSSYD FQYDATQHY LPSOGFAIP
51 DHLTGKDGK LSTHDKRKNQ KKKKKKKKK KKKKKKKKK KKKKKKKKK
101 KKEQESRTYF QOHQADGIC PTEPMHTR

11AA_SEQUENCE 1.0
ID Q9HC48 PRELIMINARY; PRT; 667 AA.

```

```

AC Q9HC48;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CTCL TUMOR ANTIGEN SE2-5 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=TESTIS;
RC MEDLINE=21143360; PubMed=1149944;
RA Fichmuller S., Usener D., Dummer R., Stein A., Thiel D.,
RA Schadenroff D.;
RT "Serological detection of cutaneous T-cell lymphoma-associated
RT antigens."
RL Proc. Natl. Acad. Sci. U.S.A. 98:629-634(2001).
DR EMBL; AF177228; AAG33676.1; -.
DR HSSP; Q12923; 3PDZ.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00585; PDZ; 2.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
FT NON_TER 1
FT NON_TER 1
SQ
SEQUENCE 667 AA; 73499 MW; C653EC168023AE02 CRC64;

Q9HC48 Length: 667 July 1, 2002 15:52 Type: P Check: 7602 ..

1 EHENVLFREND CIVRINDGL RNRREFQAH MFRQAMRTPI IMFHYPPAN
51 KEQYEQLSOS EKNNYSSRF SPDSQYIDNR SVNSAGLHY QAPRLNHP
101 EQIDSHSRPL HSAHPGKPP SAPASAPQNV FSTVSSGYN TKIKKRLNI
151 QLKKGTEGLE FSTSRDVTI GGSAPIYKN ILRGAIOQ GLKKGDRLLI
201 EVNGVDIVG SOBEVYSLR STKMEGYSL LVFRQEDAFH PRELNAEPQ
251 MQIPRETKAE DEDIVLTPDG TREPLTEVP LNDGSGAGL VSKGNSKE
301 NHADLGIFFK SIINGGAASK DGRLEVNQQL IAVNGESILG KINODAMETL
351 RRSSTEGNK RGMIOIYAR RISKCNELKS PGSPPGPELP IETALDRRR
401 RISHSLYSGI EGDIESPSRN AALSTINGES GKYQLSEPTVN MPQDVTIIE
451 DDLPLVLPFH LSDQSSSSSH DDVGFTVADA GTWAKAASID SADCSTLSPDV
501 DPVLAPOREG FGRQIADETK LNTYDDCKAG SPSPDVGPSL GLKSSSLES
551 LQTAVAEYTL NGDIPHRPR PRITGRGCN ESFFAAIDKS YKPRAVDDDD
601 EGMETLEEDT EESSRSGRES VSTASDPSH SLERQNGNQ EKGDKTRRK
651 KKKKKKKKK KKKKKKK

11AA_SEQUENCE 1.0
ID Q9H6Q7 PRELIMINARY; PRT; 720 AA.
AC Q9H6Q7;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CCNA: FLJ21979 FIS. CLONE HEP06065 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hiki J. T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Odayashi M., Nishi T., Shidohara T.,

```



RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
 RT "NEDO human cDNA sequencing project."  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK025632; BAB15196.1;  
 FT NOVELTER 720  
 SO SEQUENCE 720 AA; 84029 MW; A86586FEAA953D0B CRC64;

Q9H607 Length: 720 July 1, 2002 15:52 Type: P Check: 3351

1 MLTQVEQYT KEMKNTCIT EDLKNELQDN KGASTLSQOT HMKIQTSLDI  
 51 LKEKTKRAER TAELEADAR EKDKELVEAL KRLKDYESGV YGLEDAVVEI  
 101 KNCKNQIKIR DREIELITKE IKLELEKID FLDENEAIRE RVGLEPTWTI  
 151 DLTERNRKH LKOOQYRAEN QLLKEITESL EEBRLDKKK IRQMAQERCK  
 201 RSATSGLTTE DMLTENISQ GDRISERKLD LLSLKNMSEA QSKNEFLSRE  
 251 LIEKERDLER SRVIAKQFN KLELVEENK QLEEGMKEL QAIKEMQKDP  
 301 DVKGGETSLI IPSLERLYNA IESKNAGIF DASLHLKQV DQLTGRNEEL  
 351 RQELRESKKE AINYSQOLAK ANKIDHEK ETSLLRQSEG SNVFFKQIDL  
 401 PDGIAPSSAS IINSQNEYL IHLQLEENKE KKLKNEEDSL EDYNRKFAVI  
 451 RHQQLLYKE YLSEKETWKT ESKTIKEER KLEDOVQODA IKVKEYNNLL  
 501 NALQMSDEM KKIKAENSRK IYLYQVNEKS LIRQTTITVE LEROLKRENE  
 551 KOKNELLSME AEVCERTIGL QREKKAIFK IALQKVVON SVSLSELELA  
 601 NKQYNELTAK YRDILQKDNM LVQRTSNLEH LECENISLKE QVESINKLE  
 651 ITKEKLHIE QAMEQETKIG NESSMDKAKK SITNSDIVSI SKKKKKKKKK  
 701 KKKKKKKKK KKKKKKKKK

11AA\_SEQUENCE 1.0  
 ID Q9H5V6 PRELIMINARY; PRT; 168 AA.  
 AC Q9H5V6;  
 DT 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE CDNA: FLJ22976 FIS, CLONE KAT11222 (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,  
 RA Suzuki Y., Ohtsuka M., Nishi T., Shibahara T., Tanaka T.,  
 RA Nakamura Y., Isogai T., Sugano S.;  
 RT "NEDO human cDNA sequencing project."  
 RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK026622; BAB15513.1;  
 FT NOVELTER 168  
 SO SEQUENCE 168 AA; 19549 MW; A19BDB195F8A1A90 CRC64;

Q9H5V6 Length: 168 July 1, 2002 15:52 Type: P Check: 379

1 MNGGSSGIG QQKGNVQVA ATPASASC QYRCIECQNE AKELYRQYNH  
 51 GVLKTTICKS CQKPYDKYIE YDPIIILINA ILCKAQAYRH ILFTQJINIH  
 101 GKLCIFCLIC EAYLWMOLQ DSNONTAPDD LIRYVREMEK KKKKKKKKKK  
 151 KKKKKKKKK KKKKKKKK

11AA\_SEQUENCE 1.0  
 ID Q9NT34 PRELIMINARY; PRT; 380 AA.

AC Q9NT34;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE HYPOTHETICAL 42.7 KDA PROTEIN (FRAGMENT).  
 GN DKFZP4341120.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TESTIS;  
 RA Ottenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL137556; CAB70810.1;  
 KW Hypothetical protein  
 FT NOVELTER 380  
 SO SEQUENCE 380 AA; 42689 MW; 67F50DD101346AFB CRC64;

Q9NT34 Length: 380 July 1, 2002 15:52 Type: P Check: 7330

1 MSTDCKLNF KRAYQLTK TQPEATDA FMDQFWADTA TSYQDFVALY  
 51 PAAEIRAVRE EESPULALIC YKAVEKLYQ AESGCHSEK KOIYINCRL  
 101 LTRVLPIEIE PDWRFQFWS TVPGAGRGG EEDDEHARPL AESTLLAIAD  
 151 LTRCPDFTVQ SHRSTVDSA EDVHSLDCE YIWEAGVQFA HSPQPNYTHD  
 201 MNRMELLKLL LTCFSEAWLT PPAPEGSSTN PMVQFCFSE NNHALPLFTS  
 251 LNTVCAVDP VQYGIPIYHL LPSDYREPLV EEAQVLIYV LDHDSASSAS  
 301 PTVDGTTTGT AMDADPPGP ENLFVNYLSR IHREDFOFI LKGIARLLSN  
 351 LLLQKKKKKK KKKKKKKKK

11AA\_SEQUENCE 1.0  
 ID Q95LV6 PRELIMINARY; PRT; 531 AA.  
 AC Q95LV6;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE HYPOTHETICAL 61.4 KDA PROTEIN (FRAGMENT).  
 OS Macaca fascicularis (Orab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;  
 OC Cercopitheidae; Macaca.  
 OX NCBI\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TESTIS;  
 RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,  
 RA Terao K., Sugano S.;  
 RT "Isolation of novel full-length cDNA clones from macaque testis cDNA  
 RT libraries."  
 RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB071085; BAB64479.1;  
 KW Hypothetical protein.  
 FT NOVELTER 531  
 SO SEQUENCE 531 AA; 61389 MW; B55996B4F5CDD60C CRC64;

Q95LV6 Length: 531 July 1, 2002 15:52 Type: P Check: 7515

1 MESESSNANM MNVQEREDK NIQKMLPESV PCTSOHLSES TYQMDPPDC  
 51 KSRSEPSKPE GRSSNLSHI VQKTEQETHF RESYLEPISG YMAKOSPBNQ  
 101 EGKCVYGLK TSPPTGKSE IGSMPHDPW DENPRKWD SISEKTANWP  
 151 KNQTVLKLPL DFFSLMSSEY ESRSTLEFI GKKSITSPKH VTLKTKOLPI

```

201 SOLFNIRCS TENHKKKHQ CFKXMKGRQ WYTSIGELR SATEYAKSP
251 SKSMIDKLE NTAARGTISN RTHQNVYGH TTEEKEVQE NVAASSIGPL
301 DEFMPYLSDS KNQNTIRLS ERKTIINPKC LTMKEKSPSI SQIRKINHE
351 TRKHKKLES NLKTRKXAMW QGENVDTFF NTISFTPDS DIKQSRFOT
401 EIDMRISGS HTOPQIESL AEGIARCSK RRTSNLYKT KLHRESEK
451 KOEHLTMDP FYAENPMNT HLKRPDLG SEDVILGEF ISKQFYKGN
501 SKKKKKKKK KKKKKKKK KKKKKKKK K

11AA_SEQUENCE 1.0
ID 09LXK2: PRELIMINARY; PRT: 517 AA.
AC 09LXK2:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE HYPOTHETICAL 59.7 KDA PROTEIN.
GN T20N10.250.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
[1]
RN SEQUENCE FROM N.A.
RA D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,
RA Rued S., Lemcke K., Mayer K.F.X., Quefiet F., Salanoubat M.,
RA Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL353032; CAB8307.1;
DR InterPro: IPR001810; F-box; 1.
DR Pfam: PF00646; F-box; 1.
DR SMART: SM00256; FBOX; 1.
DR PROSITE: PS50181; FBOX; 1.
DR Hypothetical protein.
KW SEQUENCE 517 AA; 59689 MW; EC6D957D01F8E670 CRC64;
SQ
09LXK2 Length: 517 July 1, 2002 15:52 Type: P Check: 4143
1 MDLFSSLPNE LYHILSFLS TKRALTGV L SKRMNLPAF VYLEFDDSV
51 FLHPEERKRE KEGILQSFMD FVDRVLDLHG DSLIKTFSLK CKTGVDSDHV
101 DRWICNVLAR GVSDDLFD FRLYSLPHE VGVSRLLVYL RVGSSEDLXW
151 WQFLCLPML KTLVDSCLW CIGQFOILL ACALAEELDM TTRKKSNSV
201 TVSSSLIKEL TDLHGCSV VNLKSLSPDA PSIVFYCYD SLAEDYPOVN
251 LKNLVEAQIN LLTQAQIEQ VRAUNNELV ADOVFGGLN AKMLTGLN
301 VQQLYLSPT LEVLSRCSGS MPVENNLKVL SIMSDNRGW QAMPVLLRNC
351 PHETLLIEG LHYATDKG DVCDCISRDY KDHSLTSCPV KKLQIYEEFG
401 TIELEMIKH FLKIFPCKE MDIYAHENSH TLFKDPITFE RVGKKKKKKK
451 KKKKKKKKK KKKKKIRLN FKPVAKTEQF LKRLADKLCF IQQCEFLDY
501 DSSLGELALL AMDSRPS
11AA_SEQUENCE 1.0
ID 09LG29: PRELIMINARY; PRT: 260 AA.
AC 09LG29:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

```

```

DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE GENOMIC DNA, CHROMOSOME 3, BAC CLONE:FLD9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
[1]
RN SEQUENCE FROM N.A.
RA STRAIN-COLUMBIA.
RA Nakamura Y.
RT "Structural Analysis of Arabidopsis thaliana Chromosome 3. III."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF002460; BAA97098.1;
DR InterPro: IPR001365; Linker_histone.
DR PRINTS: PR00624; HISTONEH5.
SQ SEQUENCE 260 AA; 33307 MW; 43E2394CB8131143 CRC64;
09LG29 Length: 260 July 1, 2002 15:52 Type: P Check: 6094
1 MDRCIRKKK KKKKKKKK KKKKKKKK KKKKKKKK KKKKKKKK
51 KKKKKKKK KKKKKKKK KKKKKKKK KKKKKKKK KKKKKKKK
101 KKKKKKKK KKKKKKKK KKKKKKKK KKKKKKKK KKKKKKKK
151 KKKKKKKK KKKKKKKK KKKKKKKK KKKKKKKK KKKKKKKK
201 KKKKKKKK KKKKKKKK KKKKKKKK KKKKKKKK KKKKKKKK
251 INWYAGFYIF
11AA_SEQUENCE 1.0
ID 035807: PRELIMINARY; PRT: 129 AA.
AC 035807:
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE MICROVASCULAR ENDOTHELIAL DIFFERENTIATION PROTEIN 2.
GN MDG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-EPIDIDYMIS;
RX MEDLINE=98172708; PubMed=9511718;
RA Pirels F., Loser B., Marx M.;
RT "Differential expression of osteopontin, PC4, and CEC5, a novel mRNA
RT species, during in vitro angiogenesis.";
RL Exp. Cell Res. 239:1-10(1998).
DR EMBL: Y08769; CAAT0022.1;
DR InterPro: IPR000719; Euk_Pkinase.
DR Pfam: PF00069; Pkinase; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferrase.
SQ SEQUENCE 129 AA; 15080 MW; 3810227BBE2EDB4 CRC64;
035807 Length: 129 July 1, 2002 15:52 Type: P Check: 7510
1 MLKHPHVEL LETYSSDGL YWFEFMDGA DLCEIYKRA DAGFVSEAV
51 ASHWKQILE ALRYCHDNNI IHRDVPHCV LLASKKKKK KKKKKKKK
101 KKKKKKMG FDAFWAIPV KSSGVIIO
11AA_SEQUENCE 1.0
ID 064075: PRELIMINARY; PRT: 215 AA.
AC 064075:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

```

DE NUCLEOPORIN p52 HOMOLOG PROTEIN (FRAGMENT).  
OS Rattus sp. Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10118;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95151924; PubMed=7849178;  
RA Wang Z.O., Akmal K.M., Kim K.H.;  
RT "An unusual nucleoporin-related messenger ribonucleic acid is present  
in the germ cells of rat testis."  
RL Biol. Reprod. 51:102-103(1994).  
DR EMBL; S75997; AAB3384.1; .  
KW Porin.  
FT NON\_TER 1 1  
SQ SEQUENCE 215 AA; 24593 MW; 098251C97A8FBD88 CRC64;  
  
Q64075 Length: 215 July 1, 2002 15:52 Type: P Check: 8048 ..  
1 SGGRTATSCD EDCISSLPF SLGGPVKQDC EFLEKRRKK KKKKKKKKK  
51 KKKKKKTGDN AKVSROYSL KYKLEHAE QAKVELDFIL SQKLEEDLL  
101 SPLBSVKEQ SGTLYLQHAD EEREKYKLA ENIDAQLRKM AQLDKDIEH  
151 LNWAGGPAQT SDPLQIQICKI LNAHMDSLQW VDQSSALLQK RVEEASRYCE  
201 SRREQERSL RIAFD



```
1
1: (R,K){20}
   KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
2: K KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
3: KK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
4: KKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKK KKK
   (K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKK XK
   (K){20}

AAW38841 ck: 5252 len: 30 1 Aaw38841 Delivery peptide used in peptide m

1
1: (R,K){20}
   KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
2: K KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
3: KK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
4: KKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKK KKK
   (K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKK XK
   (K){20}

AAW38842 ck: 7590 len: 31 1 Aaw38842 Delivery peptide used in peptide m

1
1: (R,K){20}
   KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
2: K KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
3: KK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

1
5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKK KKK
   (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKK XK
   (K){20}

AAW38877 ck: 1129 len: 23 1 Aaw38877 Delivery peptide used in pepti.

1
1: (R,K){20}
   KKKKKKKKKKKKKKKKKKKKK VTK
   (K){20}

AAW38843 ck: 3 len: 32 1 Aaw38843 Delivery peptide used in pepti.

1
1: (R,K){20}
   KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
2: K KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
3: KK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
5: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKK
   (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKK XK
   (K){20}

AAW38878 ck: 2949 len: 24 1 Aaw38878 Delivery peptide used in pepti.

1
1: (R,K){20}
   KKKKKKKKKKKKKKKKKKKKK KVTK
   (K){20}
2: K KKKKKKKKKKKKKKKKKKKKK VTK
   (K){20}

AAW38844 ck: 2491 len: 33 1 Aaw38844 Delivery peptide used in pepti.
```

1

(R,K){20}  
 (K){20}  
 1: KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 2: K KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 3: KK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 4: KKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 5: KKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 6: KKKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 7: KKKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 8: KKKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 9: KKKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 10: KKKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 11: KKKKK KKKKKKKKKKKKKKKKK KKK

(K){20}  
 12: KKKKK KKKKKKKKKKKKKKKKK XK

AAW38879 ck: 4844 len: 25 i Aaw38879 Delivery peptide used in peptide m

1

(R,K){20}  
 (K){20}  
 1: KKKKKKKKKKKKKKKKKKK KKVTK

(K){20}  
 2: K KKKKKKKKKKKKKKKKK KVTX

(K){20}  
 3: KK KKKKKKKKKKKKKKKKK VTK

AAW38845 ck: 5054 len: 34 i Aaw38845 Delivery peptide used in peptide m

1

(R,K){20}  
 (K){20}  
 1: KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 2: K KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 3: KK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 4: KKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 5: KKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 6: KKKKK KKKKKKKKKKKKKKKKK KKKKK

1

(K){20}  
 7: KKKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 8: KKKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 9: KKKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 10: KKKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 11: KKKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 12: KKKKK KKKKKKKKKKKKKKKKK KKK

(K){20}  
 13: KKKKK KKKKKKKKKKKKKKKKK XK

AAW38880 ck: 6814 len: 26 i Aaw38880 Delivery peptide used in peptid

1

(R,K){20}  
 (K){20}  
 1: KKKKKKKKKKKKKKKKKKK KKVTK

(K){20}  
 2: K KKKKKKKKKKKKKKKKK KKVTK

(K){20}  
 3: KK KKKKKKKKKKKKKKKKK KVTX

(K){20}  
 4: KKK KKKKKKKKKKKKKKKKK VTK

AAW38846 ck: 7692 len: 35 i Aaw38846 Delivery peptide used in peptid

1

(R,K){20}  
 (K){20}  
 1: KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 2: K KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 3: KK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 4: KKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 5: KKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 6: KKKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 7: KKKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 8: KKKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 9: KKKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 10: KKKKK KKKKKKKKKKKKKKKKK KKKKK

(K){20}  
 11: KKKKK KKKKKKKKKKKKKKKKK KKKKK

```
(K){20}
12: KKKK KKKKKKKKKKKKKKK KKK
(K){20}
13: KKKK KKKKKKKKKKKKKKK KKK
(K){20}
14: KKKK KKKKKKKKKKKKKKK KKK

AAW38833 ck: 9248 len: 22 1 Aaw38833 Delivery peptide used in peptide m
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKK KKK

AAW38834 ck: 986 len: 23 1 Aaw38834 Delivery peptide used in peptide m
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKK KKK
2: K KKKKKKKKKKKKKKK KKK

AAW38835 ck: 2799 len: 24 1 Aaw38835 Delivery peptide used in peptide m
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKK KKK
2: K KKKKKKKKKKKKKKK KKK
3: K KKKKKKKKKKKKKKK KKK

AAW38836 ck: 4687 len: 25 1 Aaw38836 Delivery peptide used in peptide m
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKK KKK
2: K KKKKKKKKKKKKKKK KKK
3: K KKKKKKKKKKKKKKK KKK
4: K KKKKKKKKKKKKKKK KKK

AAW38837 ck: 6650 len: 26 1 Aaw38837 Delivery peptide used in peptide m
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKK KKK
2: K KKKKKKKKKKKKKKK KKK
3: K KKKKKKKKKKKKKKK KKK
4: K KKKKKKKKKKKKKKK KKK

(K){20}
5: KKK KKKKKKKKKKKKKKK KKK

AAW38838 ck: 8688 len: 27 1 Aaw38838 Delivery peptide used in peptide m
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKK KKK
2: K KKKKKKKKKKKKKKK KKK
3: K KKKKKKKKKKKKKKK KKK
4: K KKKKKKKKKKKKKKK KKK
5: K KKKKKKKKKKKKKKK KKK
6: K KKKKKKKKKKKKKKK KKK

AAW38796 ck: 9227 len: 22 1 Aaw38796 Delivery peptide used in peptide m
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKK KKK

AAW38797 ck: 964 len: 23 1 Aaw38797 Delivery peptide used in peptide m
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKK KKK
2: K KKKKKKKKKKKKKKK KKK

AAW38798 ck: 2776 len: 24 1 Aaw38798 Delivery peptide used in peptide m
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKK KKK
2: K KKKKKKKKKKKKKKK KKK
3: K KKKKKKKKKKKKKKK KKK

AAW38799 ck: 4663 len: 25 1 Aaw38799 Delivery peptide used in peptide m
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKK KKK
2: K KKKKKKKKKKKKKKK KKK
3: K KKKKKKKKKKKKKKK KKK
4: K KKKKKKKKKKKKKKK KKK

AAW38800 ck: 6625 len: 26 1 Aaw38800 Delivery peptide used in peptide m
(R,K){20}
```

1

```
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKK KKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKK KKKKK
```

AAW38801 ck: 8662 len: 27 1 Aaw38801 Delivery peptide used in peptide m

1

```
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKK KKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKK KKKKK
```

AAW38802 ck: 774 len: 28 1 Aaw38802 Delivery peptide used in peptide m

1

```
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKK KKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
7: KKKKK KKKKKKKKKKKKKKKKK KKKKK
```

AAW38803 ck: 2961 len: 29 1 Aaw38803 Delivery peptide used in peptide m

1

```
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKK KKKKK
```

1

```
(K){20}
3: KK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
7: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
8: KKKKK KKKKKKKKKKKKKKKKK KKKKK
```

AAW38804 ck: 5223 len: 30 1 Aaw38804 Delivery peptide used in peptide m

```
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKKKK KKKKK*
(K){20}
2: K KKKKKKKKKKKKKKKKK KKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
7: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
8: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
9: KKKKK KKKKKKKKKKKKKKKKK KKKKK
```

AAW38805 ck: 7560 len: 31 1 Aaw38805 Delivery peptide used in peptide m

1

```
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKK KKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
7: KKKKK KKKKKKKKKKKKKKKKK KKKKK
```



8: KKKK (K){20} KKKK  
9: KKKK (K){20} KKKK KKKK  
10: KKKK (K){20} KKKK KKKK

AAW38806 ck: 9972 len: 32 ! Aaw38806 Delivery peptide used in peptide π

1: (R,K){20} KKKK  
(K){20}

2: K KKKK (K){20} KKKK

3: K KKKK (K){20} KKKK

4: K KKKK (K){20} KKKK

5: KKKK (K){20} KKKK

6: KKKK (K){20} KKKK

7: KKKK (K){20} KKKK

8: KKKK (K){20} KKKK

9: KKKK (K){20} KKKK

10: KKKK (K){20} KKKK KKKK

11: KKKK (K){20} KKKK KKKK

AAW38807 ck: 2459 len: 33 ! Aaw38807 Delivery peptide used in peptide π

1: (R,K){20} KKKK  
(K){20}

2: K KKKK (K){20} KKKK

3: K KKKK (K){20} KKKK

4: K KKKK (K){20} KKKK

5: KKKK (K){20} KKKK

6: KKKK (K){20} KKKK

7: KKKK (K){20} KKKK

(K){20}

8: KKKK (K){20} KKKK

9: KKKK (K){20} KKKK KKKK

10: KKKK (K){20} KKKK KKKK

11: KKKK (K){20} KKKK KKKK

12: KKKK (K){20} KKKK KKKK

AAW38808 ck: 5021 len: 34 ! Aaw38808 Delivery peptide used in peptide π

1: (R,K){20} KKKK  
(K){20}

2: K KKKK (K){20} KKKK

3: K KKKK (K){20} KKKK

4: KKKK (K){20} KKKK

5: KKKK (K){20} KKKK

6: KKKK (K){20} KKKK

7: KKKK (K){20} KKKK

8: KKKK (K){20} KKKK

9: KKKK (K){20} KKKK

10: KKKK (K){20} KKKK

11: KKKK (K){20} KKKK KKKK

12: KKKK (K){20} KKKK KKKK

13: KKKK (K){20} KKKK KKKK

AAW38881 ck: 8859 len: 27 ! Aaw38881 Delivery peptide used in peptide π

1: (R,K){20} KKKK  
(K){20}

2: K KKKK (K){20} KKKK

3: K KKKK (K){20} KKKK

4: KKKK (K){20} KKKK

5: KKKK (K){20}  
KKKKKKKKKKKKKKKKKK VTK

AAW38847 ck: 405 len: 36 ! Aaw38847 Delivery peptide used in peptide m

1: (R,K){20}  
(K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

2: K (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

3: KK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

4: KKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

5: KKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

6: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

7: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

8: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

9: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

10: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

11: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

12: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

13: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

14: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKK

15: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKK XK

AAW38882 ck: 979 len: 28 ! Aaw38882 Delivery peptide used in peptide m

1: (R,K){20}  
(K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

2: K (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

3: KK (K){20}  
KKKKKKKKKKKKKKKKKK KKKVT

4: KKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKVT

5: KKKK (K){20}  
KKKKKKKKKKKKKKKKKK KVK

6: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK VTK

AAW38848 ck: 3193 len: 37 ! Aaw38848 Delivery peptide used in peptid

1: (R,K){20}  
(K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

2: K (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

3: KK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

4: KKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

5: KKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

6: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

7: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

8: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

9: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

10: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

11: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

12: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

13: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

14: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKK

15: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKK

16: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKK XK

AAW38883 ck: 3174 len: 29 ! Aaw38883 Delivery peptide used in peptid

1: (R,K){20}  
(K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

2: K (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

3: KK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKV

4: KKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKVT

5: KKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKVT

6: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKVT

6: KKKKK KKKKKKKKKKKKKKKKKKK KVTK

(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKK VTK

AAW38849 ck: 6056 len: 38 ! Aaw38849 Delivery peptide used in peptide m

1

(R,K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

14: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

15: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

17: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

AAW38884 ck: 5444 len: 30 ! Aaw38884 Delivery peptide used in peptide m

1

(R,K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

1

(R,K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

14: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

15: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

17: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

18: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

```
1
AAW38885 ck: 7789 len: 31 ! Aaw38885 Delivery peptide used in peptide m
(R,K){20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
7: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
8: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
9: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20} VTK

AAW38851 ck: 2007 len: 40 ! Aaw38851 Delivery peptide used in peptide m
(R,K){20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
5: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
6: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
7: KKKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
8: KKKKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
9: KKKKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
10: KKKKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
11: KKKKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
12: KKKKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
13: KKKKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
```

```
1
AAW38886 ck: 209 len: 32 ! Aaw38886 Delivery peptide used in peptid
(R,K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20} XK

AAW38852 ck: 5095 len: 41 ! Aaw38852 Delivery peptide used in peptid
(R,K){20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
7: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
8: KKKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
9: KKKKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
10: KKKKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20} VTK
```

6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

AAW38887 ck: 2704 len: 33 ! Aaw38887 Delivery peptide used in peptide m

1  
(R,K){20}  
(K){20}  
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
2: K KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

AAW38853 ck: 8258 len: 42 ! Aaw38853 Delivery peptide used in peptide m

1  
(R,K){20}  
(K){20}  
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
2: K KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
19: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
20: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

21: KKKK KKKKKKKKKKKKKKKKK XK

AAW38888 ck: 5274 len: 34 ! Aaw38888 Delivery peptide used in peptide m

(R,K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

12: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

AAW38854 ck: 1496 len: 43 ! Aaw38854 Delivery peptide used in peptide m

(R,K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

12: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

13: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

14: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

15: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

17: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

18: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

19: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

20: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

21: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

22: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

AAW38889 ck: 7919 len: 35 ! Aaw38889 Delivery peptide used in peptid

(R,K){20}

(K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

11: KKKK (K){20} KKVTK  
12: KKKK (K){20} KKVTK  
13: KKKK (K){20} VTK

AAW38890 ck: 639 len: 36 ! Aaw38890 Delivery peptide used in peptide m

1

(R,K){20}  
1: KKKK (K){20} KKKK

2: K (K){20} KKKK

3: KK (K){20} KKKK

4: KK (K){20} KKKK

5: KKKK (K){20} KKKK

6: KKKK (K){20} KKKK

7: KKKK (K){20} KKKK

8: KKKK (K){20} KKKK

9: KKKK (K){20} KKKK

10: KKKK (K){20} KKKK

11: KKKK (K){20} KKVTK

12: KKKK (K){20} KKVTK

13: KKKK (K){20} KVT

14: KKKK (K){20} VTK

AAW38891 ck: 3434 len: 37 ! Aaw38891 Delivery peptide used in peptide m

1

(R,K){20}  
1: KKKK (K){20} KKKK

2: K (K){20} KKKK

3: KK (K){20} KKKK

4: KK (K){20} KKKK

5: KKKK (K){20} KKKK

1

AAW38892 ck: 6304 len: 38 ! Aaw38892 Delivery peptide used in pepti

(R,K){20}  
1: KKKK (K){20} KKKK

2: K (K){20} KKKK

3: KK (K){20} KKKK

4: KK (K){20} KKKK

5: KKKK (K){20} KKKK

6: KKKK (K){20} KKKK

7: KKKK (K){20} KKKK

8: KKKK (K){20} KKKK

9: KKKK (K){20} KKKK

10: KKKK (K){20} KKKK

11: KKKK (K){20} KKKK

12: KKKK (K){20} KKKK

13: KKKK (K){20} KKVTK

14: KKKK (K){20} KKVTK  
 15: KKKK (K){20} KVTK  
 16: KKKK (K){20} VTK

AAW38893 ck: 9249 len: 39 1 Aaw38893 Delivery peptide used in peptide m

1: (R,K){20} KKKK

2: (K){20} K KKKK

3: (K){20} KK KKKK

4: (K){20} KKK KKKK

5: (K){20} KKKK KKKK

6: (K){20} KKKK KKKK

7: (K){20} KKKK KKKK

8: (K){20} KKKK KKKK

9: (K){20} KKKK KKKK

10: (K){20} KKKK KKKK

11: (K){20} KKKK KKKK

12: (K){20} KKKK KKKK

13: (K){20} KKKK KKKK

14: (K){20} KKKK KKKK

15: (K){20} KKKK KKVTK

16: (K){20} KKKK KVTK

17: (K){20} KKKK VTK

AAW38894 ck: 2269 len: 40 1 Aaw38894 Delivery peptide used in peptide m

1: (R,K){20} KKKK

(K){20}

2: K KKKK

3: (K){20} KK KKKK

4: (K){20} KK KKKK

5: (K){20} KKK KKKK

6: (K){20} KKKK KKKK

7: (K){20} KKKK KKKK

8: (K){20} KKKK KKKK

9: (K){20} KKKK KKKK

10: (K){20} KKKK KKKK

11: (K){20} KKKK KKKK

12: (K){20} KKKK KKKK

13: (K){20} KKKK KKKK

14: (K){20} KKKK KKKK

15: (K){20} KKKK KKVTK

16: (K){20} KKKK KVTK

17: (K){20} KKKK KVTK

18: (K){20} KKKK VTK

AAW38895 ck: 5364 len: 41 1 Aaw38895 Delivery peptide used in peptid

1: (R,K){20} KKKK

2: (K){20} K KKKK

3: (K){20} KK KKKK

4: (K){20} KKK KKKK

5: (K){20} KKKK KKKK

6: (K){20} KKKK KKKK

(K){20}



7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
AAW38896 ck: 8534 len: 42 1 Aaw38896 Delivery peptide used in peptide m  
(R,K){20}  
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
2: K KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
19: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
20: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
AAW38897 ck: 1779 len: 43 1 Aaw38897 Delivery peptide used in pepti.  
(R,K){20}  
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
2: K KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

```

14: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K){20}
15: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K){20}
16: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K){20}
17: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K){20}
18: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K){20}
19: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K){20}
20: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K){20}
21: KKKK KKKKKKKKKKKKKKKKKKK VTK
    (K){20}

AAW38898 ck: 5099 len: 44 ! Aaw38898 Delivery peptide used in peptide m
1: (R,K){20}
  (K){20}
  KKKKKKKKKKKKKKKKKKK KKKK
2: K KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
3: KK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
4: KKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
5: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
6: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
7: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
8: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
9: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
10: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
11: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
12: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
13: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
14: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
15: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}

```

```

16: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K){20}
17: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K){20}
18: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K){20}
19: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K){20}
20: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K){20}
21: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K){20}
22: KKKK KKKKKKKKKKKKKKKKKKK VTK
    (K){20}

AAW38809 ck: 7658 len: 35 ! Aaw38809 Delivery peptide used in peptid
1: (R,K){20}
  (K){20}
  KKKKKKKKKKKKKKKKKKK KKKK
2: K KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
3: KK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
4: KKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
5: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
6: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
7: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
8: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
9: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
10: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
11: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
12: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
13: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
14: KKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}

AAW38810 ck: 370 len: 36 ! Aaw38810 Delivery peptide used in peptid
1: (R,K){20}
  (K){20}
  KKKKKKKKKKKKKKKKKKK KKKK

```

1  
2: K KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
7: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
8: KKKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
9: KKKKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
10: KKKKKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
11: KKKKKKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
12: KKKKKKKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
13: KKKKKKKKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
14: KKKKKKKKKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
15: KKKKKKKKKKKKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
AAW38811 ck: 3157 len: 37 1 Aaw38811 Delivery peptide used in peptide m

1  
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}  
AAW38812 ck: 6019 len: 38 1 Aaw38812 Delivery peptide used in pepti

17: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK WK

AAW38813 ck: 8956 len: 39 ! Aaw38813 Delivery peptide used in peptide m

1: (R,K){20}  
(K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

2: K (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

3: KK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

4: KKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

5: KKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

6: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

7: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

8: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

9: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

10: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

11: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

12: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

13: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

14: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

15: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

16: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

17: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

18: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK WK

AAW38814 ck: 1968 len: 40 ! Aaw38814 Delivery peptide used in peptide m

1: (R,K){20}  
(K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

2: K (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

3: KK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

1

4: KKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

5: KKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

6: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

7: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

8: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

9: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

10: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

11: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

12: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

13: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

14: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

15: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

16: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

17: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

18: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

19: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK WK

AAW38815 ck: 5055 len: 41 ! Aaw38815 Delivery peptide used in peptid

1: (R,K){20}  
(K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

2: K (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

3: KK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

4: KKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

5: KKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

6: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

7: KKKKKK (K){20}  
KKKKKKKKKKKKKKKKKKKK KKKKK

8: KKKKK (K){20} KKKKK  
9: KKKKK (K){20} KKKKK  
10: KKKKK (K){20} KKKKK  
11: KKKKK (K){20} KKKKK  
12: KKKKK (K){20} KKKKK  
13: KKKKK (K){20} KKKKK  
14: KKKKK (K){20} KKKKK  
15: KKKKK (K){20} KKKKK  
16: KKKKK (K){20} KKKKK  
17: KKKKK (K){20} KKKKK  
18: KKKKK (K){20} KKKKK  
19: KKKKK (K){20} KKKKK  
20: KKKKK (K){20} KKKKK WK  
AAM38816 ck: 8217 len: 42 1 Aaw38816 Delivery peptide used in peptide m  
(R,K){20}  
1: KKKKK (K){20} KKKKK  
2: K KKKKK (K){20} KKKKK  
3: KK KKKKK (K){20} KKKKK  
4: KKK KKKKK (K){20} KKKKK  
5: KKKK KKKKK (K){20} KKKKK  
6: KKKKK KKKKK (K){20} KKKKK  
7: KKKKK KKKKK (K){20} KKKKK  
8: KKKKK KKKKK (K){20} KKKKK  
9: KKKKK KKKKK (K){20} KKKKK  
10: KKKKK KKKKK (K){20} KKKKK

11: KKKKK (K){20} KKKKK  
12: KKKKK (K){20} KKKKK  
13: KKKKK (K){20} KKKKK  
14: KKKKK (K){20} KKKKK  
15: KKKKK (K){20} KKKKK  
16: KKKKK (K){20} KKKKK  
17: KKKKK (K){20} KKKKK  
18: KKKKK (K){20} KKKKK  
19: KKKKK (K){20} KKKKK  
20: KKKKK (K){20} KKKKK WK  
21: KKKKK (K){20} KKKKK WK  
AAM38817 ck: 1454 len: 43 1 Aaw38817 Delivery peptide used in peptic  
(R,K){20}  
1: KKKKK (K){20} KKKKK  
2: K KKKKK (K){20} KKKKK  
3: KK KKKKK (K){20} KKKKK  
4: KKK KKKKK (K){20} KKKKK  
5: KKKK KKKKK (K){20} KKKKK  
6: KKKKK KKKKK (K){20} KKKKK  
7: KKKKK KKKKK (K){20} KKKKK  
8: KKKKK KKKKK (K){20} KKKKK  
9: KKKKK KKKKK (K){20} KKKKK  
10: KKKKK KKKKK (K){20} KKKKK  
11: KKKKK KKKKK (K){20} KKKKK  
12: KKKKK KKKKK (K){20} KKKKK

(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
AAW24865 ck: 2211 len: 40 ! Aaw24865 Bifunctional peptide I for binding  
(R,K){20}  
21: YEDES KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
1  
AAW24450 ck: 8137 len: 45 ! Aaw24450 Nucleic acid (NA) binding peptide  
(R,K){20}  
4: YKA KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
5: YKAK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
6: YKAKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
7: KAKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
8: AKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
AAW21590 ck: 4875 len: 30 ! Aaw21590 Antibiotic potentiating peptide  
(R,K){20}  
1: KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
2: K KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
3: KK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
4: KKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
5: KKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
6: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
AAW21591 ck: 5075 len: 434 ! Aaw21591 Antibiotic potentiating peptide  
(R,K){20}

[illegible]

25:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
26:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
27:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
28:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
29:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
30:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
31:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
32:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
33:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
34:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
35:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
36:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
37:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
38:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
39:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
40:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
41:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
42:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
43:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
44:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
45:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
46:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
47:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
48:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX
49:	XXXXX	(X)120	XXXXXXXXXXXXXXXXXXXX





98: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
99: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
100: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
101: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
102: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
103: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
104: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
105: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
106: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
107: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
108: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
109: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
110: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
111: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
112: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
113: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
114: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
115: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
116: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
117: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
118: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
119: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
120: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
121: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
122: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

123: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
124: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
125: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
126: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
127: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
128: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
129: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
130: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
131: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
132: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
133: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
134: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
135: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
136: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
137: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
138: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
139: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
140: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
141: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
142: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
143: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
144: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
145: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
146: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

147: KKKK (K){20} KKKK  
 148: KKKK (K){20} KKKK  
 149: KKKK (K){20} KKKK  
 150: KKKK (K){20} KKKK  
 151: KKKK (K){20} KKKK  
 152: KKKK (K){20} KKKK  
 153: KKKK (K){20} KKKK  
 154: KKKK (K){20} KKKK  
 155: KKKK (K){20} KKKK  
 156: KKKK (K){20} KKKK  
 157: KKKK (K){20} KKKK  
 158: KKKK (K){20} KKKK  
 159: KKKK (K){20} KKKK  
 160: KKKK (K){20} KKKK  
 161: KKKK (K){20} KKKK  
 162: KKKK (K){20} KKKK  
 163: KKKK (K){20} KKKK  
 164: KKKK (K){20} KKKK  
 165: KKKK (K){20} KKKK  
 166: KKKK (K){20} KKKK  
 167: KKKK (K){20} KKKK  
 168: KKKK (K){20} KKKK  
 169: KKKK (K){20} KKKK  
 170: KKKK (K){20} KKKK

171: KKKK KKKK KKKK KKKK KKKK  
 172: KKKK (K){20} KKKK  
 173: KKKK (K){20} KKKK  
 174: KKKK (K){20} KKKK  
 175: KKKK (K){20} KKKK  
 176: KKKK (K){20} KKKK  
 177: KKKK (K){20} KKKK  
 178: KKKK (K){20} KKKK  
 179: KKKK (K){20} KKKK  
 180: KKKK (K){20} KKKK  
 181: KKKK (K){20} KKKK  
 182: KKKK (K){20} KKKK  
 183: KKKK (K){20} KKKK  
 184: KKKK (K){20} KKKK  
 185: KKKK (K){20} KKKK  
 186: KKKK (K){20} KKKK  
 187: KKKK (K){20} KKKK  
 188: KKKK (K){20} KKKK  
 189: KKKK (K){20} KKKK  
 190: KKKK (K){20} KKKK  
 191: KKKK (K){20} KKKK  
 192: KKKK (K){20} KKKK  
 193: KKKK (K){20} KKKK  
 194: KKKK (K){20} KKKK  
 195: KKKK (K){20} KKKK

196: KKKK (K){20} KKKK  
197: KKKK (K){20} KKKK  
198: KKKK (K){20} KKKK  
199: KKKK (K){20} KKKK  
200: KKKK (K){20} KKKK  
201: KKKK (K){20} KKKK  
202: KKKK (K){20} KKKK  
203: KKKK (K){20} KKKK  
204: KKKK (K){20} KKKK  
205: KKKK (K){20} KKKK  
206: KKKK (K){20} KKKK  
207: KKKK (K){20} KKKK  
208: KKKK (K){20} KKKK  
209: KKKK (K){20} KKKK  
210: KKKK (K){20} KKKK  
211: KKKK (K){20} KKKK  
212: KKKK (K){20} KKKK  
213: KKKK (K){20} KKKK  
214: KKKK (K){20} KKKK  
215: KKKK (K){20} KKKK  
216: KKKK (K){20} KKKK  
217: KKKK (K){20} KKKK  
218: KKKK (K){20} KKKK  
219: KKKK (K){20} KKKK

9

220: KKKK (K){20} KKKK  
221: KKKK (K){20} KKKK  
222: KKKK (K){20} KKKK  
223: KKKK (K){20} KKKK  
224: KKKK (K){20} KKKK  
225: KKKK (K){20} KKKK  
226: KKKK (K){20} KKKK  
227: KKKK (K){20} KKKK  
228: KKKK (K){20} KKKK  
229: KKKK (K){20} KKKK  
230: KKKK (K){20} KKKK  
231: KKKK (K){20} KKKK  
232: KKKK (K){20} KKKK  
233: KKKK (K){20} KKKK  
234: KKKK (K){20} KKKK  
235: KKKK (K){20} KKKK  
236: KKKK (K){20} KKKK  
237: KKKK (K){20} KKKK  
238: KKKK (K){20} KKKK  
239: KKKK (K){20} KKKK  
240: KKKK (K){20} KKKK  
241: KKKK (K){20} KKKK  
242: KKKK (K){20} KKKK  
243: KKKK (K){20} KKKK

244: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 245: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 246: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 247: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 248: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 249: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 250: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 251: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 252: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 253: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 254: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 255: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 256: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 257: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 258: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 259: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 260: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 261: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 262: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 263: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 264: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 265: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 266: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 267: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 268: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}

269: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 270: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 271: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 272: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 273: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 274: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 275: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 276: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 277: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 278: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 279: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 280: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 281: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 282: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 283: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 284: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 285: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 286: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 287: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 288: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 289: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 290: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 291: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}  
 292: KKKK KKKKKKKKKKKKKKKKKKK  
 (K){20}

293: KKKK (K){20} KKKK  
294: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
295: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
296: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
297: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
298: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
299: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
300: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
301: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
302: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
303: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
304: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
305: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
306: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
307: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
308: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
309: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
310: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
311: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
312: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
313: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
314: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
315: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
316: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK

317: KKKK KKKKKKKKKKKKKKKKK KKKK  
318: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
319: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
320: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
321: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
322: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
323: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
324: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
325: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
326: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
327: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
328: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
329: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
330: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
331: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
332: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
333: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
334: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
335: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
336: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
337: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
338: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
339: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
340: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK  
341: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK

342: KKKK (K){20} KKKK  
 343: KKKK (K){20} KKKK  
 344: KKKK (K){20} KKKK  
 345: KKKK (K){20} KKKK  
 346: KKKK (K){20} KKKK  
 347: KKKK (K){20} KKKK  
 348: KKKK (K){20} KKKK  
 349: KKKK (K){20} KKKK  
 350: KKKK (K){20} KKKK  
 351: KKKK (K){20} KKKK  
 352: KKKK (K){20} KKKK  
 353: KKKK (K){20} KKKK  
 354: KKKK (K){20} KKKK  
 355: KKKK (K){20} KKKK  
 356: KKKK (K){20} KKKK  
 357: KKKK (K){20} KKKK  
 358: KKKK (K){20} KKKK  
 359: KKKK (K){20} KKKK  
 360: KKKK (K){20} KKKK  
 361: KKKK (K){20} KKKK  
 362: KKKK (K){20} KKKK  
 363: KKKK (K){20} KKKK  
 364: KKKK (K){20} KKKK  
 365: KKKK (K){20} KKKK

366: KKKK (K){20} KKKK  
 367: KKKK (K){20} KKKK  
 368: KKKK (K){20} KKKK  
 369: KKKK (K){20} KKKK  
 370: KKKK (K){20} KKKK  
 371: KKKK (K){20} KKKK  
 372: KKKK (K){20} KKKK  
 373: KKKK (K){20} KKKK  
 374: KKKK (K){20} KKKK  
 375: KKKK (K){20} KKKK  
 376: KKKK (K){20} KKKK  
 377: KKKK (K){20} KKKK  
 378: KKKK (K){20} KKKK  
 379: KKKK (K){20} KKKK  
 380: KKKK (K){20} KKKK  
 381: KKKK (K){20} KKKK  
 382: KKKK (K){20} KKKK  
 383: KKKK (K){20} KKKK  
 384: KKKK (K){20} KKKK  
 385: KKKK (K){20} KKKK  
 386: KKKK (K){20} KKKK  
 387: KKKK (K){20} KKKK  
 388: KKKK (K){20} KKKK  
 389: KKKK (K){20} KKKK

390: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
391: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
392: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
393: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
394: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
395: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
396: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
397: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
398: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
399: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
400: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
401: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
402: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
403: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
404: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
405: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
406: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
407: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
408: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
409: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
410: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
411: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
412: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
413: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
414: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

415: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
AAW65939 ck: 1569 len: 40 1 Aaw65939 Polyllysine peptide NBC32. 11/15  
1  
(R,K){20}  
2: T KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
3: TK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
4: TRK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
5: TKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
6: TKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
AAW48808 ck: 9658 len: 56 1 Aaw48808 Homo sapiens clone CG109\_1 prot  
1  
(R,K){20}  
34: EEPRE KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
35: EPREK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
36: FREKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

```

37: REKKK KKKKKKKKKKKKKKKKKK
AAM45801 ck: 9500 len: 39 1 Aa45801 One chain of a bombesin dimer. 6/1
      (R,K){20}
      (K){20}
2: C KKKKKKKKKKKKKKKKKKK XGGGQ
AAM45800 ck: 9056 len: 39 1 Aa45800 One chain of an alpha-melanocyte s
      (R,K){20}
      (K){20}
18: VGGGX KKKKKKKKKKKKKKKKKK CX
AAM45802 ck: 7664 len: 35 1 Aa45802 One chain of an alpha-MSH receptor
      (R,K){20}
      (K){20}
14: VGGGX KKKKKKKKKKKKKKKKKK CX
AAY43246 ck: 9752 len: 32 1 Aay43246 Cell-surface molecule binding pep
      (R,K){20}
      (K){20}
12: SGSGS KKKKKKKKKKKKKKKKKK K
      (K){20}
13: GSGSK KKKKKKKKKKKKKKKKKK
AAY07213 ck: 2211 len: 40 1 Aay07213 Peptide transfection vector #1. 7/
      (R,K){20}
      (K){20}
21: YEDSS KKKKKKKKKKKKKKKKKK
AAY12950 ck: 5821 len: 62 1 Aay12950 Amino acid sequence of a human seq
      (R,K){20}
      (K){20}
40: QFOAS KKKKKKKKKKKKKKKKKK KKK
      (K){20}
41: FOASK KKKKKKKKKKKKKKKKKK KK
      (K){20}
42: QASKK KKKKKKKKKKKKKKKKKK K
      (K){20}
43: ASKKK KKKKKKKKKKKKKKKKKK
AAB59105 ck: 8456 len: 27 1 Aab59105 Breast and ovarian cancer associat
      (R,K){20}
      (K){20}
6: NSAXX KKKKKKKKKKKKKKKKKK KK
      (K){20}
7: SAXXX KKKKKKKKKKKKKKKKKK K
      (K){20}
8: AXKKK KKKKKKKKKKKKKKKKKK
AAB53249 ck: 4945 len: 59 1 Aab53249 Human colon cancer antigen protein
      (R,K){20}

```

```

27: ALSLK KKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
28: LLSLK KKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
29: LSLKK KKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
30: SLKKK KKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
31: LKKKK KKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKK KKKKK
AAB53659 ck: 3850 len: 184 1 Aab53659 Human colon cancer antigen prot
      (R,K){20}
      (K){20}
1: KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
2: K KKKKKKKKKKKKKKKKK KKKKK
      (K){20}
3: KK KKKKKKKKKKKKKKKKK KKKKK
      (K){20}
4: KKK KKKKKKKKKKKKKKKKK KKKKK
      (K){20}
5: KKKK KKKKKKKKKKKKKKKKK KKKKK
      (K){20}
6: KKKKK KKKKKKKKKKKKKKKKK KKKKK
AAB53800 ck: 296 len: 69 1 Aab53800 Human colon cancer antigen prot
      (R,K){20}
      (K){20}
20: NYDCE KKKKKKKKKKKKKKKKK KKKKK
      (K){20}
21: VDCEK KKKKKKKKKKKKKKKKK KKKKK
      (K){20}
22: DCEKK KKKKKKKKKKKKKKKKK KKKKK
      (K){20}
23: CEKKK KKKKKKKKKKKKKKKKK KKKKK
      (K){20}
24: EKKKK KKKKKKKKKKKKKKKKK KKKKK
      (K){20}
25: KKKKK KKKKKKKKKKKKKKKKK KKKKK
      (K){20}
26: KKKKK KKKKKKKKKKKKKKKKK KKKKK
      (K){20}
27: KKKKK KKKKKKKKKKKKKKKKK KKKKK
      (K){20}
28: KKKKK KKKKKKKKKKKKKKKKK KKKKK
      (K){20}
29: KKKKK KKKKKKKKKKKKKKKKK KKKKK

```



```

30: KKKK (K){20}
31: KKKK (K){20}
32: KKKK (K){20}
33: KKKK (K){20}
34: KKKK (K){20}
35: KKKK (K){20}
36: KKKK (K){20}
37: KKKK (K){20}
38: KKKK (K){20}
39: KKKK (K){20}
40: KKKK (K){20}

AAB53806 ck: 8373 len: 64 ! Aab53806 Human colon cancer antigen protein
(R,K){20}
25: IDCS KKKK (K){20}
26: DCDSK KKKK (K){20}
27: CDSKK KKKK (K){20}
28: DSXKK KKKK (K){20}
29: SKKKK KKKK (K){20}
30: KKKKK KKKK (K){20}

AAB53977 ck: 7756 len: 75 ! Aab53977 Human colon cancer antigen protein
(R,K){20}
24: QFCLL KKKK (K){20}
25: FCLLK KKKK (K){20}
26: CLKKK KKKK (K){20}
27: LLKKK KKKK (K){20}
1
```

```

28: LKKKK KKKK
29: KKKKK KKKK (K){20}
30: KKKKK KKKK (K){20}
31: KKKKK KKKK (K){20}
32: KKKKK KKKK (K){20}
33: KKKKK KKKK (K){20}
34: KKKKK KKKK (K){20}
35: KKKKK KKKK (K){20}
36: KKKKK KKKK (K){20}
37: KKKKK KKKK (K){20}
38: KKKKK KKKK (K){20}
39: KKKKK KKKK (K){20}
40: KKKKK KKKK (K){20}
41: KKKKK KKKK (K){20}
42: KKKKK KKKK (K){20}
43: KKKKK KKKK (K){20}
44: KKKKK KKKK (K){20}
45: KKKKK KKKK (K){20}
46: KKKKK KKKK (K){20}
47: KKKKK KKKK (K){20}
48: KKKKK KKKK (K){20}
49: KKKKK KKKK (K){20}

AAB53980 ck: 881 len: 45 ! Aab53980 Human colon cancer antigen pro
(R,K){20}
8: ICSFQ KKKK (K){20}
1
```

9: CSFQK KKKKKKKKKKKKKKKKKKK KKKXG  
 (K){20}  
 10: SFQK KKKKKKKKKKKKKKKKKKK KXXG  
 (K){20}  
 11: FQK KKKKKKKKKKKKKKKKKKK KXGR  
 (K){20}  
 12: QK KKKKKKKKKKKKKKKKKKK XGRF

AAB54314 ck: 6235 len: 55 1 Aab54314 Human pancreatic cancer antigen pr

(R,K){20}

(K){20}

24: FFXK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

25: FXXK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

26: XXXK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

27: XXXK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

28: XK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

29: KKKK KKKKKKKKKKKKKKKKKKK KKKG

(K){20}

30: KKKK KKKKKKKKKKKKKKKKKKK KKGK

(K){20}

31: KKKK KKKKKKKKKKKKKKKKKKK KGGF

(K){20}

32: KKKK KKKKKKKKKKKKKKKKKKK GGRF

AAB56121 ck: 5941 len: 125 1 Aab56121 Human secreted protein sequence en

(R,K){20}

(K){20}

87: ILKSQ KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

88: LKSQ KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

89: KSQK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

90: SQK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

91: QKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

92: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

93: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

94: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

95: KKKK KKKKKKKKKKKKKKKKKKK KKKK

96: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

97: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

98: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

99: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

100: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

101: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

102: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

103: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

AAB27956 ck: 6732 len: 139 1 Aab27956 Human secreted protein SEQ ID N

(R,K){20}

(K){20}

91: HPRG KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

92: PRG KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

93: PRG KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

94: PRG KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

95: GRG KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

96: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

97: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

98: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

99: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

100: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

101: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

102: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

103: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

104: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

105: KKKK KKKKKKKKKKKKKKKKKKK KKKK

106: KKKKK (K){20} KKKKK  
107: KKKKK (K){20} KKKKK  
108: KKKKK (K){20} KKKKK  
109: KKKKK (K){20} KKKKK  
110: KKKKK (K){20} KKKKK  
111: KKKKK (K){20} KKKKK  
112: KKKKK (K){20} KKKKK  
113: KKKKK (K){20} KKKKK  
114: KKKKK (K){20} KKKKK  
115: KKKKK (K){20} KKKKK  
116: KKKKK (K){20} KKKKK  
117: KKKKK (K){20} KKKKK  
118: KKKKK (K){20} KKKKK  
119: KKKKK (K){20} KKKKK  
120: KKKKK (K){20} KKKKK  
AAB39140 ck: 9531 len: 66 1 Aab39140 Human secreted protein #48. 2/2001  
(R,K){20}  
41: MWTVX KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
42: WTVXK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
43: TVXKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
44: VXXKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
45: XKKKK KKKKKKKKKKKKKKKKKKK KK  
(K){20}  
46: KKKKK KKKKKKKKKKKKKKKKKKK K  
(K){20}  
47: KKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
AAB41457 ck: 4536 len: 168 1 Aab41457 Human ORFX ORF1221 polypeptide seq

1  
144: KQOKP KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
145: QOKPK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
146: QKPKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
147: KPKKK KKKKKKKKKKKKKKKKKKK KK  
(K){20}  
148: PKKKK KKKKKKKKKKKKKKKKKKK K  
(K){20}  
149: KKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
AAB42786 ck: 4959 len: 102 1 Aab42786 Human ORFX ORF2550 polypeptide  
(R,K){20}  
81: SDVLQ KKKKKKKKKKKKKKKKKKK KK  
(K){20}  
82: DVLQK KKKKKKKKKKKKKKKKKKK K  
(K){20}  
83: VLOKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
AAB42889 ck: 6432 len: 62 1 Aab42889 Human ORFX ORF2653 polypeptide  
(R,K){20}  
40: SHRCL KKKKKKKKKKKKKKKKKKK KKK  
(K){20}  
41: HRCLK KKKKKKKKKKKKKKKKKKK KK  
(K){20}  
42: RCLCK KKKKKKKKKKKKKKKKKKK K  
(K){20}  
43: CLKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
AAB43641 ck: 5216 len: 133 1 Aab43641 Human cancer associated protei  
(R,K){20}  
114: GHEQS KKKKKKKKKKKKKKKKKKK  
(K){20}  
AAB43835 ck: 4025 len: 223 1 Aab43835 Human cancer associated protei  
(R,K){20}  
196: NILFW KKKKKKKKKKKKKKKKKKK XXGGA  
(K){20}  
AAB44188 ck: 4991 len: 43 1 Aab44188 Human cancer associated protei  
(R,K){20}  
2: E KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
3: EK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

```

4:  EKK KKKKKKKKKKKKKKKKKKKKK KKKGR
    (K){20}
5:  EKKK KKKKKKKKKKKKKKKKKKKKK KKGSP
    (K){20}
6:  EKKKK KKKKKKKKKKKKKKKKKKKKK KGSPF
    (K){20}
7:  KKKKK KKKKKKKKKKKKKKKKKKKKK GPPFX
    (K){20}

AAB44380  ck: 8490  len: 42  ! Aab44380 Human secreted protein encoded by
    (R,K){20}
22:  IWKRI KKKKKKKKKKKKKKKKKKKKK K
    (K){20}
23:  WKRIK KKKKKKKKKKKKKKKKKKKKK
    (K){20}

AAB43331  ck: 8131  len: 66  ! Aab43331 Human secreted protein sequence en
    (R,K){20}
40:  IHMKP KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
41:  HHMKP KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
42:  HKPKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
43:  KPKKK KKKKKKKKKKKKKKKKKKKKK KKKK
    (K){20}
44:  PKKKK KKKKKKKKKKKKKKKKKKKKK KKK
    (K){20}
45:  KKKKK KKKKKKKKKKKKKKKKKKKKK KK
    (K){20}
46:  KKKKK KKKKKKKKKKKKKKKKKKKKK K
    (K){20}
47:  KKKKK KKKKKKKKKKKKKKKKKKKKK
    (K){20}

AAB23585  ck: 9893  len: 36  ! Aab23585 Ask21 linker peptide. 1/2001
    (R,K){20}
15:  PGSGS KKKKKKKKKKKKKKKKKKKKK KG
    (K){20}
16:  GSGSK KKKKKKKKKKKKKKKKKKKKK G
    (K){20}

AAB23586  ck: 58  len: 36  ! Aab23586 Gsk21 linker peptide. 1/2001
    (R,K){20}
15:  SGSGS KKKKKKKKKKKKKKKKKKKKK KG
    (K){20}
16:  GSGSK KKKKKKKKKKKKKKKKKKKKK G
    (K){20}

AAB23591  ck: 7684  len: 630  ! Aab23591 Modified fibre protein encoded in
    (R,K){20}

```

---

```

596:  PGSGS KKKKKKKKKKKKKKKKKKKKK KGSYS
    (K){20}
597:  GSGSK KKKKKKKKKKKKKKKKKKKKK GSYSM
    (K){20}

AAB23592  ck: 7647  len: 630  ! Aab23592 Modified fibre protein encoded
    (R,K){20}
596:  SGSGS KKKKKKKKKKKKKKKKKKKKK KGSYS
    (K){20}
597:  GSGSK KKKKKKKKKKKKKKKKKKKKK GSYSM
    (K){20}

AAB23593  ck: 2249  len: 640  ! Aab23593 Modified fibre protein encoded
    (R,K){20}
596:  PGSGS KKKKKKKKKKKKKKKKKKKKK KGSAL
    (K){20}
597:  GSGSK KKKKKKKKKKKKKKKKKKKKK GSAEK
    (K){20}

AAB23594  ck: 2612  len: 640  ! Aab23594 Modified fibre protein encoded
    (R,K){20}
596:  SGSGS KKKKKKKKKKKKKKKKKKKKK KGSAL
    (K){20}
597:  GSGSK KKKKKKKKKKKKKKKKKKKKK GSAEK
    (K){20}

AAB13780  ck: 7317  len: 21  ! Aab13780 Soluble peptide antigen pk. 11/
    (R,K){20}
2:  C KKKKKKKKKKKKKKKKKKKKKKKKKKK
    (K){20}

AAB13783  ck: 4553  len: 45  ! Aab13783 Soluble tandem pEA/ PK peptide
    (R,K){20}
26:  AAAAA KKKKKKKKKKKKKKKKKKKKK
    (K){20}

AAB13784  ck: 4126  len: 44  ! Aab13784 Soluble tandem HA/ PK peptide c
    (R,K){20}
25:  DGYTG KKKKKKKKKKKKKKKKKKKKK
    (K){20}

AAG00834  ck: 6330  len: 103  ! Aag00834 Human secreted protein, SEQ ID
    (R,K){20}
83:  RKROQ KKKKKKKKKKKKKKKKKKKKK K
    (K){20}
84:  KRQOK KKKKKKKKKKKKKKKKKKKKK
    (K){20}

AAY98493  ck: 8137  len: 45  ! Aay98493 Peptide #5 used in nucleic acid
    (R,K){20}
4:  YKA KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}

```

5: YKAK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
6: YKAK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
7: KAKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
8: AKKKK (K){20} KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

AAy98495 ck: 4361 len: 59 1 Aay98495 Nuclear ligand used in nucleic aci  
(R,K){20}  
18: APYKA KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
19: PYKAK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
20: YKAKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

21: KAKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
22: AKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
27: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
28: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
29: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
30: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
31: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
32: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
33: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
34: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
35: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
36: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
37: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
38: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

AAy98497 ck: 4925 len: 100 1 Aay98497 Peptide #8 used in nucleic aci  
(R,K){20}  
1: KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
2: K KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
3: KK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
4: KKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
5: KKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}

6: KKKK (K){20} KKKK  
7: KKKK (K){20} KKKK  
8: KKKK (K){20} KKKK  
9: KKKK (K){20} KKKK  
10: KKKK (K){20} KKKK  
11: KKKK (K){20} KKKK  
12: KKKK (K){20} KKKK  
13: KKKK (K){20} KKKK  
14: KKKK (K){20} KKKK  
15: KKKK (K){20} KKKK  
16: KKKK (K){20} KKKK  
17: KKKK (K){20} KKKK  
18: KKKK (K){20} KKKK  
19: KKKK (K){20} KKKK  
20: KKKK (K){20} KKKK  
21: KKKK (K){20} KKKK  
22: KKKK (K){20} KKKK  
23: KKKK (K){20} KKKK  
24: KKKK (K){20} KKKK  
25: KKKK (K){20} KKKK  
26: KKKK (K){20} KKKK  
27: KKKK (K){20} KKKK  
28: KKKK (K){20} KKKK  
29: KKKK (K){20} KKKK

30: KKKK (K){20} KKKK  
31: KKKK (K){20} KKKK  
32: KKKK (K){20} KKKK  
33: KKKK (K){20} KKKK  
34: KKKK (K){20} KKKK  
35: KKKK (K){20} KKKK  
36: KKKK (K){20} KKKK  
37: KKKK (K){20} KKKK  
38: KKKK (K){20} KKKK  
39: KKKK (K){20} KKKK  
40: KKKK (K){20} KKKK  
41: KKKK (K){20} KKKK  
42: KKKK (K){20} KKKK  
43: KKKK (K){20} KKKK  
44: KKKK (K){20} KKKK  
45: KKKK (K){20} KKKK  
46: KKKK (K){20} KKKK  
47: KKKK (K){20} KKKK  
48: KKKK (K){20} KKKK  
49: KKKK (K){20} KKKK  
50: KKKK (K){20} KKKK  
51: KKKK (K){20} KKKK  
52: KKKK (K){20} KKKK  
53: KKKK (K){20} KKKK

54: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
55: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
56: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
57: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
58: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
59: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
60: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
61: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
62: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
63: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
64: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
65: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
66: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
67: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
68: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
69: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
70: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
71: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
72: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
73: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
74: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
75: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
76: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
77: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
78: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

79: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
80: KKKK KKKKKKKKKKKKKKKKKKK K  
(K){20}  
81: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
AAY81805 cx: 1885 len: 351 1 Aay81805 Murine mahogany protein sequen  
(R,K){20}  
(R,K){20}  
3: EF KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
4: EFR KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
5: EFRK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
6: EFRK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
7: FRKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
8: RKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
9: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
10: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
11: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
12: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
13: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
14: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
15: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
16: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
17: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
18: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
19: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
20: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
21: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
22: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

23: KKKK (K){20} KKKK  
24: KKKK (K){20} KKKK  
25: KKKK (K){20} KKKK  
26: KKKK (K){20} KKKK  
27: KKKK (K){20} KKKK

AAY56902 ck: 4875 len: 30 1 Aay56902 (Lys)30 protein sequence. 4/2000

1

1: - (R,K){20} KKKK  
(K){20}

2: K (K){20} KKKK

3: KK (K){20} KKKK

4: KKK (K){20} KKKK

5: KKKK (K){20} KKKK

6: KKKK (K){20} KKKK

7: KKKK (K){20} KKKK

8: KKKK (K){20} KKK

9: KKKK (K){20} KK

10: KKKK (K){20} K

11: KKKK (K){20} KKKK

AAY56903 ck: 5075 len: 434 1 Aay56903 (Lys)434 protein sequence. 4/2000

1

1: (R,K){20} KKKK  
(K){20}

2: K (K){20} KKKK

3: KK (K){20} KKKK

4: KKK (K){20} KKKK

5: KKKK (K){20} KKKK

(K){20}

6: KKKK (K){20} KKKK  
7: KKKK (K){20} KKKK  
8: KKKK (K){20} KKKK  
9: KKKK (K){20} KKKK  
10: KKKK (K){20} KKKK  
11: KKKK (K){20} KKKK  
12: KKKK (K){20} KKKK  
13: KKKK (K){20} KKKK  
14: KKKK (K){20} KKKK  
15: KKKK (K){20} KKKK  
16: KKKK (K){20} KKKK  
17: KKKK (K){20} KKKK  
18: KKKK (K){20} KKKK  
19: KKKK (K){20} KKKK  
20: KKKK (K){20} KKKK  
21: KKKK (K){20} KKKK  
22: KKKK (K){20} KKKK  
23: KKKK (K){20} KKKK  
24: KKKK (K){20} KKKK  
25: KKKK (K){20} KKKK  
26: KKKK (K){20} KKKK  
27: KKKK (K){20} KKKK  
28: KKKK (K){20} KKKK  
29: KKKK (K){20} KKKK  
30: KKKK (K){20} KKKK



31: KKKK (K){20} KKKK  
32: KKKK (K){20} KKKK  
33: KKKK (K){20} KKKK  
34: KKKK (K){20} KKKK  
35: KKKK (K){20} KKKK  
36: KKKK (K){20} KKKK  
37: KKKK (K){20} KKKK  
38: KKKK (K){20} KKKK  
39: KKKK (K){20} KKKK  
40: KKKK (K){20} KKKK  
41: KKKK (K){20} KKKK  
42: KKKK (K){20} KKKK  
43: KKKK (K){20} KKKK  
44: KKKK (K){20} KKKK  
45: KKKK (K){20} KKKK  
46: KKKK (K){20} KKKK  
47: KKKK (K){20} KKKK  
48: KKKK (K){20} KKKK  
49: KKKK (K){20} KKKK  
50: KKKK (K){20} KKKK  
51: KKKK (K){20} KKKK  
52: KKKK (K){20} KKKK  
53: KKKK (K){20} KKKK  
54: KKKK (K){20} KKKK

55: KKKK (K){20} KKKK  
56: KKKK (K){20} KKKK  
57: KKKK (K){20} KKKK  
58: KKKK (K){20} KKKK  
59: KKKK (K){20} KKKK  
60: KKKK (K){20} KKKK  
61: KKKK (K){20} KKKK  
62: KKKK (K){20} KKKK  
63: KKKK (K){20} KKKK  
64: KKKK (K){20} KKKK  
65: KKKK (K){20} KKKK  
66: KKKK (K){20} KKKK  
67: KKKK (K){20} KKKK  
68: KKKK (K){20} KKKK  
69: KKKK (K){20} KKKK  
70: KKKK (K){20} KKKK  
71: KKKK (K){20} KKKK  
72: KKKK (K){20} KKKK  
73: KKKK (K){20} KKKK  
74: KKKK (K){20} KKKK  
75: KKKK (K){20} KKKK  
76: KKKK (K){20} KKKK  
77: KKKK (K){20} KKKK  
78: KKKK (K){20} KKKK

79: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 80: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 81: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 82: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 83: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 84: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 85: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 86: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 87: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 88: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 89: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 90: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 91: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 92: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 93: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 94: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 95: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 96: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 97: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 98: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 99: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 100: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 101: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 102: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 103: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}

104: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 105: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 106: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 107: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 108: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 109: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 110: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 111: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 112: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 113: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 114: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 115: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 116: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 117: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 118: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 119: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 120: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 121: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 122: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 123: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 124: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 125: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 126: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}  
 127: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
 (K){20}

128: KKKK (K){20} KKKK  
129: KKKK (K){20} KKKK  
130: KKKK (K){20} KKKK  
131: KKKK (K){20} KKKK  
132: KKKK (K){20} KKKK  
133: KKKK (K){20} KKKK  
134: KKKK (K){20} KKKK  
135: KKKK (K){20} KKKK  
136: KKKK (K){20} KKKK  
137: KKKK (K){20} KKKK  
138: KKKK (K){20} KKKK  
139: KKKK (K){20} KKKK  
140: KKKK (K){20} KKKK  
141: KKKK (K){20} KKKK  
142: KKKK (K){20} KKKK  
143: KKKK (K){20} KKKK  
144: KKKK (K){20} KKKK  
145: KKKK (K){20} KKKK  
146: KKKK (K){20} KKKK  
147: KKKK (K){20} KKKK  
148: KKKK (K){20} KKKK  
149: KKKK (K){20} KKKK  
150: KKKK (K){20} KKKK  
151: KKKK (K){20} KKKK

152: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
153: KKKK (K){20} KKKK  
154: KKKK (K){20} KKKK  
155: KKKK (K){20} KKKK  
156: KKKK (K){20} KKKK  
157: KKKK (K){20} KKKK  
158: KKKK (K){20} KKKK  
159: KKKK (K){20} KKKK  
160: KKKK (K){20} KKKK  
161: KKKK (K){20} KKKK  
162: KKKK (K){20} KKKK  
163: KKKK (K){20} KKKK  
164: KKKK (K){20} KKKK  
165: KKKK (K){20} KKKK  
166: KKKK (K){20} KKKK  
167: KKKK (K){20} KKKK  
168: KKKK (K){20} KKKK  
169: KKKK (K){20} KKKK  
170: KKKK (K){20} KKKK  
171: KKKK (K){20} KKKK  
172: KKKK (K){20} KKKK  
173: KKKK (K){20} KKKK  
174: KKKK (K){20} KKKK  
175: KKKK (K){20} KKKK  
176: KKKK (K){20} KKKK

177: KKKKK (K){20} KKKKK  
 178: KKKKK (K){20} KKKKK  
 179: KKKKK (K){20} KKKKK  
 180: KKKKK (K){20} KKKKK  
 181: KKKKK (K){20} KKKKK  
 182: KKKKK (K){20} KKKKK  
 183: KKKKK (K){20} KKKKK  
 184: KKKKK (K){20} KKKKK  
 185: KKKKK (K){20} KKKKK  
 186: KKKKK (K){20} KKKKK  
 187: KKKKK (K){20} KKKKK  
 188: KKKKK (K){20} KKKKK  
 189: KKKKK (K){20} KKKKK  
 190: KKKKK (K){20} KKKKK  
 191: KKKKK (K){20} KKKKK  
 192: KKKKK (K){20} KKKKK  
 193: KKKKK (K){20} KKKKK  
 194: KKKKK (K){20} KKKKK  
 195: KKKKK (K){20} KKKKK  
 196: KKKKK (K){20} KKKKK  
 197: KKKKK (K){20} KKKKK  
 198: KKKKK (K){20} KKKKK  
 199: KKKKK (K){20} KKKKK  
 200: KKKKK (K){20} KKKKK



201: KKKKK (K){20} KKKKK  
 202: KKKKK (K){20} KKKKK  
 203: KKKKK (K){20} KKKKK  
 204: KKKKK (K){20} KKKKK  
 205: KKKKK (K){20} KKKKK  
 206: KKKKK (K){20} KKKKK  
 207: KKKKK (K){20} KKKKK  
 208: KKKKK (K){20} KKKKK  
 209: KKKKK (K){20} KKKKK  
 210: KKKKK (K){20} KKKKK  
 211: KKKKK (K){20} KKKKK  
 212: KKKKK (K){20} KKKKK  
 213: KKKKK (K){20} KKKKK  
 214: KKKKK (K){20} KKKKK  
 215: KKKKK (K){20} KKKKK  
 216: KKKKK (K){20} KKKKK  
 217: KKKKK (K){20} KKKKK  
 218: KKKKK (K){20} KKKKK  
 219: KKKKK (K){20} KKKKK  
 220: KKKKK (K){20} KKKKK  
 221: KKKKK (K){20} KKKKK  
 222: KKKKK (K){20} KKKKK  
 223: KKKKK (K){20} KKKKK  
 224: KKKKK (K){20} KKKKK

225: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
226: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
227: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
228: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
229: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
230: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
231: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
232: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
233: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
234: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
235: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
236: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
237: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
238: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
239: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
240: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
241: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
242: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
243: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
244: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
245: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
246: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
247: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
248: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
249: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}

250: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
251: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
252: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
253: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
254: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
255: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
256: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
257: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
258: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
259: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
260: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
261: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
262: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
263: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
264: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
265: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
266: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
267: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
268: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
269: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
270: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
271: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
272: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}  
273: KKKK KKKKKKKKKKKKKKKKK KKKK  
      (K){20}

274: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 275: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 276: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 277: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 278: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 279: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 280: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 281: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 282: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 283: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 284: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 285: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 286: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 287: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 288: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 289: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 290: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 291: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 292: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 293: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 294: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 295: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 296: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 297: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK

298: KKKK KKKKKKKKKKKKKKKKKKK KKKK  
 299: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 300: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 301: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 302: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 303: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 304: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 305: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 306: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 307: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 308: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 309: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 310: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 311: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 312: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 313: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 314: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 315: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 316: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 317: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 318: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 319: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 320: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 321: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK  
 322: KKKK (K) [20] KKKKKKKKKKKKKKKKKKK KKKK

323: KKKK (K){20} KKKK  
324: KKKK (K){20} KKKK  
325: KKKK (K){20} KKKK  
326: KKKK (K){20} KKKK  
327: KKKK (K){20} KKKK  
328: KKKK (K){20} KKKK  
329: KKKK (K){20} KKKK  
330: KKKK (K){20} KKKK  
331: KKKK (K){20} KKKK  
332: KKKK (K){20} KKKK  
333: KKKK (K){20} KKKK  
334: KKKK (K){20} KKKK  
335: KKKK (K){20} KKKK  
336: KKKK (K){20} KKKK  
337: KKKK (K){20} KKKK  
338: KKKK (K){20} KKKK  
339: KKKK (K){20} KKKK  
340: KKKK (K){20} KKKK  
341: KKKK (K){20} KKKK  
342: KKKK (K){20} KKKK  
343: KKKK (K){20} KKKK  
344: KKKK (K){20} KKKK  
345: KKKK (K){20} KKKK  
346: KKKK (K){20} KKKK

347: KKKK (K){20} KKKK  
348: KKKK (K){20} KKKK  
349: KKKK (K){20} KKKK  
350: KKKK (K){20} KKKK  
351: KKKK (K){20} KKKK  
352: KKKK (K){20} KKKK  
353: KKKK (K){20} KKKK  
354: KKKK (K){20} KKKK  
355: KKKK (K){20} KKKK  
356: KKKK (K){20} KKKK  
357: KKKK (K){20} KKKK  
358: KKKK (K){20} KKKK  
359: KKKK (K){20} KKKK  
360: KKKK (K){20} KKKK  
361: KKKK (K){20} KKKK  
362: KKKK (K){20} KKKK  
363: KKKK (K){20} KKKK  
364: KKKK (K){20} KKKK  
365: KKKK (K){20} KKKK  
366: KKKK (K){20} KKKK  
367: KKKK (K){20} KKKK  
368: KKKK (K){20} KKKK  
369: KKKK (K){20} KKKK  
370: KKKK (K){20} KKKK

371: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
372: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
373: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
374: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
375: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
376: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
377: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
378: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
379: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
380: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
381: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
382: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
383: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
384: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
385: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
386: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
387: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
388: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
389: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
390: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
391: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
392: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
393: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
394: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
395: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

1  
396: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
397: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
398: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
399: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
400: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
401: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
402: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
403: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
404: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
405: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
406: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
407: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
408: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
409: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
410: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
411: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
412: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
413: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
414: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
415: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

AY86248 ck: 8783 len: 128 : Aay86248 Human secreted protein HCHPF68,  
(R,K){20}  
84: ANPP KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
85: NPPEP KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
86: PPEPK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}



```

87: PPKKK KKKKKKKKKKKKKKKKKKK KKKK      (K){20}
88: PKKKK KKKKKKKKKKKKKKKKKKK KKKK      (K){20}
89: KKKKK KKKKKKKKKKKKKKKKKKK KKKK      (K){20}
90: KKKKK KKKKKKKKKKKKKKKKKKK KKKK      (K){20}
91: KKKKK KKKKKKKKKKKKKKKKKKK KKKK      (K){20}
92: KKKKK KKKKKKKKKKKKKKKKKKK KKKK      (K){20}
93: KKKKK KKKKKKKKKKKKKKKKKKK KKKK      (K){20}
94: KKKKK KKKKKKKKKKKKKKKKKKK KKKK      (K){20}
95: KKKKK KKKKKKKKKKKKKKKKKKK KKKK      (K){20}
96: KKKKK KKKKKKKKKKKKKKKKKKK KKKK      (K){20}
97: KKKKK KKKKKKKKKKKKKKKKKKK KKKK      (K){20}
98: KKKKK KKKKKKKKKKKKKKKKKKK KKKK      (K){20}
99: KKKKK KKKKKKKKKKKKKKKKKKK KKKK      (K){20}
100: KKKKK KKKKKKKKKKKKKKKKKKK KKKK      (K){20}
101: KKKKK KKKKKKKKKKKKKKKKKKK KKKK      (K){20}
102: KKKKK KKKKKKKKKKKKKKKKKKK KKKK      (K){20}
103: KKKKK KKKKKKKKKKKKKKKKKKK KKKK      (K){20}
104: KKKKK KKKKKKKKKKKKKKKKKKK KKKK      (K){20}
105: KKKKK KKKKKKKKKKKKKKKKKKK KKKK      (K){20}
106: KKKKK KKKKKKKKKKKKKKKKKKK KKKK      (K){20}

AAV59038 ck: 8137 len: 45 ! Aay59038 Peptide used in the construction c
(R,K){20}
(K){20}
4: YKA KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
5: YKAK KKKKKKKKKKKKKKKKK KKKK
(K){20}
6: YKAK KKKKKKKKKKKKKKKKK KKKK
(K){20}

```

```

7: KAKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
8: AKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
9: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
10: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
11: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
12: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
13: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
14: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
15: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
16: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
17: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
18: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
19: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
20: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
21: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
22: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
23: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
24: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

AAy59040 ck: 4361 len: 59 1 Aay59040 Nuclear ligand for transportin
(R,K){20}
(K){20}
18: APYKA KKKKKKKKKKKKKKKKKKK KKKKK
19: PYKAK KKKKKKKKKKKKKKKKKKK KKKKK
20: YKAKK KKKKKKKKKKKKKKKKKKK KKKKK
21: KAKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
22: AKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

```

23: KKKK (K){20} KKKK  
 24: KKKK (K){20} KKKK  
 25: KKKK (K){20} KKKK  
 26: KKKK (K){20} KKKK  
 27: KKKK (K){20} KKKK  
 28: KKKK (K){20} KKKK  
 29: KKKK (K){20} KKKK  
 30: KKKK (K){20} KKKK  
 31: KKKK (K){20} KKKK  
 32: KKKK (K){20} KKKK  
 33: KKKK (K){20} KKKK  
 34: KKKK (K){20} KKKK  
 35: KKKK (K){20} KKKK  
 36: KKKK (K){20} KKKK  
 37: KKKK (K){20} KKKK  
 38: KKKK (K){20} KKKK

AA59042 ck: 4925 len: 100 1 Aay59042 amino acid polymer seg ID NO: 62 C

1: (R,K){20} KKKK  
 2: K (K){20} KKKK  
 3: KK (K){20} KKKK  
 4: KKK (K){20} KKKK  
 5: KKKK (K){20} KKKK  
 6: KKKK (K){20} KKKK  
 7: KKKK (K){20} KKKK

8: KKKK (K){20} KKKK  
 9: KKKK (K){20} KKKK  
 10: KKKK (K){20} KKKK  
 11: KKKK (K){20} KKKK  
 12: KKKK (K){20} KKKK  
 13: KKKK (K){20} KKKK  
 14: KKKK (K){20} KKKK  
 15: KKKK (K){20} KKKK  
 16: KKKK (K){20} KKKK  
 17: KKKK (K){20} KKKK  
 18: KKKK (K){20} KKKK  
 19: KKKK (K){20} KKKK  
 20: KKKK (K){20} KKKK  
 21: KKKK (K){20} KKKK  
 22: KKKK (K){20} KKKK  
 23: KKKK (K){20} KKKK  
 24: KKKK (K){20} KKKK  
 25: KKKK (K){20} KKKK  
 26: KKKK (K){20} KKKK  
 27: KKKK (K){20} KKKK  
 28: KKKK (K){20} KKKK  
 29: KKKK (K){20} KKKK  
 30: KKKK (K){20} KKKK  
 31: KKKK (K){20} KKKK

32: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
33: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
34: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
35: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
36: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
37: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
38: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
39: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
40: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
41: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
42: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
43: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
44: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
45: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
46: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
47: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
48: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
49: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
50: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
51: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
52: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
53: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
54: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
55: KKKK (K)[20]  
KKKKKKKKKKKKKKKKKK KKKK  
(K)[20]

56: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
57: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
58: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
59: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
60: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
61: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
62: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
63: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
64: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
65: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
66: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
67: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
68: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
69: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
70: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
71: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
72: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
73: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
74: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
75: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
76: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
77: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K)[20]  
78: KKKK KKKKKKKKKKKKKKKKK KKK  
(K)[20]  
79: KKKK KKKKKKKKKKKKKKKKK KK  
(K)[20]  
80: KKKK KKKKKKKKKKKKKKKKK K  
(K)[20]

81: KKKK (K){20}  
KKKKKKKKKKKKKKKK

ABG00401 ck: 5116 len: 1,074 ! Abg00401 Novel human diagnostic protein #38

(R,K){20}  
609: RGSSS KKKRRKKKKKKKKRR KRRK  
(R,K){20}

610: GSSSK KKKRRKKKKKKRRR NRKK  
(R,K){20}

ABG03974 ck: 5341 len: 99 ! Abg03974 Novel human diagnostic protein #35

1 (R,K){20}  
(K){20}

2: M KKKKKKKKKKKKKKKKK

3: MK KKKKKKKKKKKKKKKK  
(K){20}

4: MK KKKKKKKKKKKKKKKK  
(K){20}

5: MKK KKKKKKKKKKKKKK KNSQI  
(K){20}

6: MKKK KKKKKKKKKKKKK NSQID  
(K){20}

ABG04391 ck: 9047 len: 139 ! Abg04391 Novel human diagnostic protein #43

1 (R,K){20}  
(K){20}

79: EEEEE KKKKKKKKKKKKKK

80: EEEK (K){20}  
KKKKKKKKKKKKKKK

81: EEEK (K){20}  
KKKKKKKKKKKKKKK

82: EEEK (K){20}  
KKKKKKKKKKKKKKK

83: EKKK (K){20}  
KKKKKKKKKKKKKKK

84: KKKK (K){20}  
KKKKKKKKKKKKKKK

85: KKKK (K){20}  
KKKKKKKKKKKKKKK

86: KKKK (K){20}  
KKKKKKKKKKKKKKK

87: KKKK (K){20}  
KKKKKKKKKKKKKKK

88: KKKK (K){20}  
KKKKKKKKKKKKKKK

89: KKKK (K){20}  
KKKKKKKKKKKKKKK

90: KKKK (K){20}  
KKKKKKKKKKKKKKK

91: KKKK (K){20}  
KKKKKKKKKKKKKKK

92: KKKK (K){20}  
KKKKKKKKKKKKKKK

93: KKKK (K){20}  
KKKKKKKKKKKKKKK

94: KKKK (K){20}  
KKKKKKKKKKKKKKK

95: KKKK (K){20}  
KKKKKKKKKKKKKKK

96: KKKK (K){20}  
KKKKKKKKKKKKKKK

97: KKKK (K){20}  
KKKKKKKKKKKKKKK

98: KKKK (K){20}  
KKKKKKKKKKKKKKK

99: KKKK (K){20}  
KKKKKKKKKKKKKKK

100: KKKK (K){20}  
KKKKKKKKKKKKKKK

101: KKKK (K){20}  
KKKKKKKKKKKKKKK

102: KKKK (K){20}  
KKKKKKKKKKKKKKK

103: KKKK (K){20}  
KKKKKKKKKKKKKKK

104: KKKK (K){20}  
KKKKKKKKKKKKKKK

105: KKKK (K){20}  
KKKKKKKKKKKKKKK

106: KKKK (K){20}  
KKKKKKKKKKKKKKK

107: KKKK (K){20}  
KKKKKKKKKKKKKKK

108: KKKK (K){20}  
KKKKKKKKKKKKKKK

109: KKKK (K){20}  
KKKKKKKKKKKKKKK

110: KKKK (K){20}  
KKKKKKKKKKKKKKK

111: KKKK (K){20}  
KKKKKKKKKKKKKKK

112: KKKK (K){20}  
KKKKKKKKKKKKKKK

113: KKKK (K){20}  
KKKKKKKKKKKKKKK

ABG05352 ck: 1276 len: 204 ! Abg05352 Novel human diagnostic protein

106: EEEEE (K) [20]  
KKKKKKKKKKKKKKKKKKKK  
107: EEEEE (K) [20]  
KKKKKKKKKKKKKKKKKKKK  
108: EEEEE (K) [20]  
KKKKKKKKKKKKKKKKKKKK  
109: EEEEE (K) [20]  
KKKKKKKKKKKKKKKKKKKK  
110: EEEEE (K) [20]  
KKKKKKKKKKKKKKKKKKKK  
111: KKKKK (K) [20]  
KKKKKKKKKKKKKKKKKKKK  
112: KKKKK (K) [20]  
KKKKKKKKKKKKKKKKKKKK  
113: KKKKK (R,K) [20]  
KKKKKKKKKKKKKKKKKKKK  
114: KKKKK (R,K) [20]  
KKKKKKKKKKKKKKKKKKKK  
115: KKKKK (R,K) [20]  
KKKKKKKKKKKKKKKKKKKK  
116: KKKKK (R,K) [20]  
KKKKKKKKKKKKKKKKKKKK  
117: KKKKK (R,K) [20]  
KKKKKKKKKKKKKKKKKKKK  
118: KKKKK (R,K) [20]  
KKKKKKKKKKKKKKKKKKKK  
ABG05367 ck: 6907 len: 808 1 Abg05367 Novel human diagnostic protein #53  
1  
219: EEEEE (R,K) [20]  
RRKKKKKKKKKKKKKKKKKK  
220: EEEEE (R,K) [20]  
RRKKKKKKKKKKKKKKKKKK  
221: EEEEE (R,K) [20]  
RRKKKKKKKKKKKKKKKKKK  
222: EEEEE (R,K) [20]  
RRKKKKKKKKKKKKKKKKKK  
223: EEEEE (R,K) [20]  
RRKKKKKKKKKKKKKKKKKK  
224: RKKKK (R,K) [20]  
RRKKKKKKKKKKKKKKKKKK  
225: RKKKK (R,K) [20]  
RRKKKKKKKKKKKKKKKKKK  
226: KKKKK (R,K) [20]  
RRKKKKKKKKKKKKKKKKKK  
227: KKKKK (R,K) [20]  
RRKKKKKKKKKKKKKKKKKK  
228: KKKKK (R,K) [20]  
RRKKKKKKKKKKKKKKKKKK

1  
229: KKKKK (R,K) [20]  
RRRRRRRRRRRRRRRRRRRR  
230: KKKKK (R,K) [20]  
RRRRRRRRRRRRRRRRRRRR  
231: KKKKK (R,K) [20]  
RRRRRRRRRRRRRRRRRRRR  
232: KKKKK (R,K) [20]  
RRRRRRRRRRRRRRRRRRRR  
233: KKKKK (R,K) [20]  
RRRRRRRRRRRRRRRRRRRR  
234: RRRRR (R,K) [20]  
RRRRRRRRRRRRRRRRRRRR  
235: RRRRR (R,K) [20]  
RRRRRRRRRRRRRRRRRRRR  
236: RRRRR (R,K) [20]  
RRRRRRRRRRRRRRRRRRRR  
237: RRRRR (R,K) [20]  
RRRRRRRRRRRRRRRRRRRR  
238: KKKKK (K) [20]  
KKKKKKKKKKKKKKKKKKKK  
239: KKKKK (K) [20]  
KKKKKKKKKKKKKKKKKKKK  
240: KKKKK (K) [20]  
KKKKKKKKKKKKKKKKKKKK  
ABG06375 ck: 7807 len: 2,570 1 Abg06375 Novel human diagnostic protein  
1  
205: KKKSS (R,K) [20]  
RRRRRRRRRRRRRRRRRRRR  
206: KKKSS (R,K) [20]  
RRRRRRRRRRRRRRRRRRRR  
207: KSSRR (R,K) [20]  
RRRRRRRRRRRRRRRRRRRR  
208: SSRRR (R,K) [20]  
RRRRRRRRRRRRRRRRRRRR  
ABG06513 ck: 2934 len: 154 1 Abg06513 Novel human diagnostic protein  
1  
49: KKKKT (R,K) [20]  
RRKKKKKKKKKKKKKKKKKK  
50: KKKTR (R,K) [20]  
RRKKKKKKKKKKKKKKKKKK  
51: KKTTR (R,K) [20]  
RRKKKKKKKKKKKKKKKKKK  
52: KTRKK (R,K) [20]  
RRKKKKKKKKKKKKKKKKKK  
53: TRKKR (R,K) [20]  
RRKKKKKKKKKKKKKKKKKK  
54: RKKRK (R,K) [20]  
RRKKKKKKKKKKKKKKKKKK

55: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
56: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
57: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
58: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
59: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
60: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
61: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
62: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
63: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
64: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
65: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
ABG07742 ck: 8672 len: 502 ! Abg07742 Novel human diagnostic protein #77  
54: DDEEE (R,K){20} RRRKKKKKKKKKKKKKKKKKK KKKKK  
55: DEEER (R,K){20} RRRKKKKKKKKKKKKKKKKKK KKKKK  
56: DEEER (R,K){20} RRRKKKKKKKKKKKKKKKKKK KKKKK  
57: DEEER (R,K){20} RRRKKKKKKKKKKKKKKKKKK KKKKK  
58: DEEER (R,K){20} RRRKKKKKKKKKKKKKKKKKK KKKKK  
59: DEEER (R,K){20} RRRKKKKKKKKKKKKKKKKKK KKKKK  
60: DEEER (R,K){20} RRRKKKKKKKKKKKKKKKKKK KKKKK  
61: DEEER (R,K){20} RRRKKKKKKKKKKKKKKKKKK KKKKK  
62: DEEER (R,K){20} RRRKKKKKKKKKKKKKKKKKK KKKKK  
63: DEEER (R,K){20} RRRKKKKKKKKKKKKKKKKKK KKKKK  
64: DEEER (R,K){20} RRRKKKKKKKKKKKKKKKKKK KKKKK  
65: DEEER (R,K){20} RRRKKKKKKKKKKKKKKKKKK KKKKK

66: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
67: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK

ABG10052 ck: 7107 len: 39 ! Abg10052 Novel human diagnostic protein

2: Q (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
3: Q (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
4: Q (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
5: Q (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
6: Q (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
7: Q (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK

ABG10053 ck: 3274 len: 189 ! Abg10053 Novel human diagnostic protein

70: EKEKE (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
71: EKEKE (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
72: EKEKE (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
73: EKEKE (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
74: EKEKE (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
75: EKEKE (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
76: EKEKE (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
77: EKEKE (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
78: EKEKE (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
79: EKEKE (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
ABG11241 ck: 3870 len: 121 ! Abg11241 Novel human diagnostic protein  
44: KKKKE (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
45: KKKKE (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK

46: KEEKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKE  
47: EEEKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
48: EKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKE  
49: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

ABG11242 ck: 4493 len: 100 ! Abg11242 Novel human diagnostic protein #11

1  
(R,K){20}

42: ETPSE (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

43: TPSEK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

44: PSEKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

45: SEKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

46: EKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

47: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

48: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

49: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

50: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

51: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

52: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

53: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

54: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

55: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

56: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

57: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

58: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

59: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

60: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

1

61: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

62: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

63: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

64: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

65: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

66: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

67: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

68: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

69: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

70: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

71: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

ABG11245 ck: 2517 len: 85 ! Abg11245 Novel human diagnostic protein

(R,K){20}

33: EEEBE (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

34: EEEBK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

35: EEEBK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

36: EEEKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

37: EKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

38: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

39: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

40: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

41: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

42: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

43: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

44: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

45: KKKK (K){20} KKKK  
46: KKKK (K){20} KKKK  
47: KKKK (K){20} KKKK  
48: KKKK (K){20} KKKK  
49: KKKK (K){20} KKKK  
50: KKKK (K){20} KKKK  
51: KKKK (K){20} KKKK  
52: KKKK (K){20} KKKK  
ABG11250 ck: 4343 len: 92 ! Abg11250 Novel human diagnostic protein #11  
(R,K){20}  
53: EKEE (R,K){20} KKKK  
(K){20}  
54: KEKE (K){20} KKKK  
(K){20}  
55: EKEE (K){20} KKKK  
(K){20}  
56: KEKE (K){20} KKKK  
(R,K){20}  
57: EKEE (R,K){20} KKKK  
(R,K){20}  
58: KKKK (R,K){20} KKKK  
(R,K){20}  
59: KKKK (R,K){20} KKKK  
ABG11266 ck: 9563 len: 146 ! Abg11266 Novel human diagnostic protein #11  
(R,K){20}  
36: KEKE (K){20} KKKK  
(K){20}  
37: EKEE (K){20} KKKK  
(K){20}  
38: KEKE (K){20} KKKK  
(K){20}  
39: REKE (K){20} KKKK  
(K){20}  
40: EKKK (K){20} EEEE  
(R,K){20}  
79: EEEE (R,K){20} KKKK  
(R,K){20}  
80: EEEE (R,K){20} KKKK

81: EEEE (R,K){20} KKKK  
(R,K){20}  
82: EEEE (R,K){20} KKKK  
(R,K){20}  
83: EKKK (R,K){20} KKKK  
(R,K){20}  
84: KKKK (R,K){20} KKKK  
(R,K){20}  
85: KKKK (R,K){20} KKKK  
(R,K){20}  
86: KKKK (R,K){20} KKKK  
(R,K){20}  
87: KKKK (R,K){20} KKKK  
(R,K){20}  
88: KKKK (R,K){20} KKKK  
(R,K){20}  
89: KKKK (R,K){20} KKKK  
(R,K){20}  
90: KKKK (R,K){20} KKKK  
(R,K){20}  
91: KKKK (R,K){20} KKKK  
(R,K){20}  
92: KKKK (R,K){20} KKKK  
(R,K){20}  
93: KKKK (R,K){20} KKKK  
ABG11277 ck: 8026 len: 1,080 ! Abg11277 Novel human diagnostic protein  
(R,K){20}  
709: KEKE (R,K){20} KKKK  
(R,K){20}  
710: KEKE (R,K){20} KKKK  
(R,K){20}  
711: EKEE (R,K){20} KKKK  
(R,K){20}  
712: KEKE (R,K){20} KKKK  
(R,K){20}  
713: EKEE (R,K){20} KKKK  
(R,K){20}  
714: KKKK (R,K){20} KKKK  
(R,K){20}  
715: KKKK (R,K){20} KKKK  
(R,K){20}  
716: KKKK (R,K){20} KKKK  
(R,K){20}  
717: KKKK (R,K){20} KKKK  
(R,K){20}  
718: KKKK (R,K){20} KKKK





(R,K){20}  
228: RERER RRRRRRRRRRRRRRRR RRRGX  
(R,K){20}  
229: ERERR RRRRRRRRRRRRRRRR RRGXE  
(R,K){20}  
230: RERRR RRRRRRRRRRRRRRRR RGXEF  
(R,K){20}  
231: ERRRR RRRRRRRRRRRRRRRR GXEFL

ABG26213 ck: 6773 len: 735 1 Abg26213 Novel human diagnostic protein #26  
(R,K){20}  
(R,K){20}  
173: RGSSS KKKRRRRRRRRRRRRR KNRRK  
(R,K){20}  
174: GSSSS KKKRRRRRRRRRRRRR NRKKK

1  
ABG26488 ck: 523 len: 124 1 Abg26488 Novel human diagnostic protein #26  
(R,K){20}  
(R,K){20}  
91: EEEEE RRRRRRRRRRRRRRRR RKKKK  
(R,K){20}  
92: EEEER RRRRRRRRRRRRRRRR KKKKK  
(R,K){20}  
93: EEERR RRRRRRRRRRRRRRRR KKKKR  
(R,K){20}  
94: EERRR RRRRRRRRRRRRRRRR KKKRK  
(R,K){20}  
95: ERRRR RRRRRRRRRRRRRRRR KKKKK  
(R,K){20}  
96: RRRRR KKKRRRRRRRRRRRRR KRRKK  
(R,K){20}  
97: RRRRK KKKRRRRRRRRRRRRR RKKKK  
(R,K){20}  
98: RRRKK KKKRRRRRRRRRRRRR KKKKK  
(R,K){20}  
99: RRRKK KKKRRRRRRRRRRRRR KKKKK  
(R,K){20}  
100: RKKKK KKKRRRRRRRRRRRRR KKKKE  
(R,K){20}  
101: KKKKK KKKRRRRRRRRRRRRR KKKK  
(R,K){20}  
102: KKKKK KKKRRRRRRRRRRRRR KKK  
(R,K){20}  
103: KKKKK KKKRRRRRRRRRRRRR KK  
(R,K){20}  
104: KKKKK KKKRRRRRRRRRRRRR E  
ABG26489 ck: 9923 len: 120 1 Abg26489 Novel human diagnostic protein #26  
(R,K){20}

(R){20}  
70: RRRRG RRRRRRRRRRRRRRRR RRRKK  
(R){20}  
71: RRRGR RRRRRRRRRRRRRRRR RKKKE  
(R){20}  
72: RRGRR RRRRRRRRRRRRRRRR KKKKE  
(R,K){20}  
73: RGRRR RRRRRRRRRRRRRRRR KKEEE  
(R,K){20}  
74: GRRRR RRRRRRRRRRRRRRRR KEEEE  
(R,K){20}  
75: RRRRR RRRRRRRRRRRRRRRR EEEEE

1  
ABG26490 ck: 4146 len: 96 1 Abg26490 Novel human diagnostic protein  
(R,K){20}  
(R){20}  
39: KKGEE RRRRRRRRRRRRRRRR RRRGR  
(R){20}  
40: KGEE RRRRRRRRRRRRRRRR RRGGR  
(R){20}  
41: GEEER RRRRRRRRRRRRRRRR RGGRR  
(R){20}  
42: EERRR RRRRRRRRRRRRRRRR GGGGG  
(R){20}  
73: GGGGG RRRRRRRRRRRRRRRR RRRR  
(R){20}  
74: RGGRR RRRRRRRRRRRRRRRR RRR  
(R){20}  
75: GGGRR RRRRRRRRRRRRRRRR RR  
(R){20}  
76: RGGRR RRRRRRRRRRRRRRRR R  
(R){20}  
77: GRRRR RRRRRRRRRRRRRRRR  
ABG26491 ck: 8179 len: 109 1 Abg26491 Novel human diagnostic protein  
(R,K){20}  
(R,K){20}  
75: RRRGE KKKRRRRRRRRRRRRR KKKKK  
(R,K){20}  
76: RRGKK KKKRRRRRRRRRRRRR KKKKK  
(R,K){20}  
77: RGEKK KKKRRRRRRRRRRRRR KKKKK  
(R,K){20}  
78: GEKKK KKKRRRRRRRRRRRRR KKKKK  
(R,K){20}  
79: EKKKK KKKRRRRRRRRRRRRR KKKKK  
(R){20}  
80: KKKKR KKKRRRRRRRRRRRRR KKKGN  
(R){20}  
81: KKKRK KKKRRRRRRRRRRRRR KKGNT

```
(K){20}
82: KRRKK KKKKKKKKKKKKKKKK KGNLS
      (K){20}
83: KKKKK KKKKKKKKKKKKKKKK GNLSV

ABG26492  ck: 5234  len: 68      ! Abg26492 Novel human diagnostic protein #26
      (R,K){20}
29: KKKEE RRRRRRRRRRRRRRRRRR RRRRG
      (R){20}
30: KKEER RRRRRRRRRRRRRRRRRR RRRGK
      (R){20}
31: KEERR RRRRRRRRRRRRRRRRRR RRGKK
      (R){20}
32: EERRR RRRRRRRRRRRRRRRRRR RGKGD
      (R){20}
33: ERRRR RRRRRRRRRRRRRRRRRR GKLDG

ABG26493  ck: 4204  len: 80      ! Abg26493 Novel human diagnostic protein #26
      (R,K){20}
      (R,K){20}
27: EKEKE KRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
28: KEKEK RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
29: EKEKR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
30: KEKRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
31: EKRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
32: KRRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
33: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
34: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
35: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
36: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
37: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
38: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
39: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
40: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
```

```
41: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
42: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
43: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
44: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
45: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
46: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
47: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
48: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
49: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
50: RRRRR RRRRRRRRRRRRRRRRRR RRRRT
      (R){20}
51: RRRRR RRRRRRRRRRRRRRRRRR RRRTN
      (R){20}
52: RRRRR RRRRRRRRRRRRRRRRRR RRRTN
      (R){20}
53: RRRRR RRRRRRRRRRRRRRRRRR RRRNN

ABG26496  ck: 8829  len: 90      ! Abg26496 Novel human diagnostic protein
      (R,K){20}
      (R,K){20}
61: EKKEE KRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
62: KKEEK RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
63: KEEKR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
64: EEKRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
65: EKRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
66: KRRRR RRRRRRRRRRRRRRRRRR RRRRR
      (R){20}
67: RRRRR RRRRRRRRRRRRRRRRRR RRRNN
      (R){20}
68: RRRRR RRRRRRRRRRRRRRRRRR RRRN
      (R){20}
69: RRRRR RRRRRRRRRRRRRRRRRR RRRN
      (R){20}
70: RRRRR RRRRRRRRRRRRRRRRRR RRRN
```



72: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
73: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
74: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
75: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
76: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
77: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
78: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
79: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
80: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
81: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
82: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
83: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
84: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
85: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
86: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
ABG26498 ck: 7156 len: 140 1 Abg26498 Novel human diagnostic protein #26  
(R,K)(20)  
97: RGGG RRRRRRRRRRRRRRRRRRR (R)(20)  
98: RGGG RRRRRRRRRRRRRRRRRRR (R)(20)  
99: GGGR RRRRRRRRRRRRRRRRRRR (R)(20)  
100: GGRR RRRRRRRRRRRRRRRRRRR (R)(20)  
101: GRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
102: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
103: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
104: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)

105: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
106: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
107: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
108: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
109: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
110: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
111: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
112: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
113: RRRR RRRRRRRRRRRRRRRRRRR (R,K)(20)  
ABG26500 ck: 7252 len: 78 1 Abg26500 Novel human diagnostic protein  
(R,K)(20)  
35: GGEE RRRRRRRRRRRRRRRRRRR (R)(20)  
36: GEEER RRRRRRRRRRRRRRRRRRR (R)(20)  
37: EEER RRRRRRRRRRRRRRRRRRR (R)(20)  
38: EERR RRRRRRRRRRRRRRRRRRR (R)(20)  
39: EERR RRRRRRRRRRRRRRRRRRR (R)(20)  
40: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
41: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
42: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
43: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
44: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
45: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
46: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
47: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)  
48: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)

1

ABG26501 ck: 1730 len: 182 1 Abg26501 Novel human diagnostic protein #26

```
(R,K)(20)
72: KEEKE KKKKKKKKKKKKKKKKKKK KKKKE
      (K)(20)
73: EEKEE KKKKKKKKKKKKKKKKKKK KKKKE
      (K)(20)
74: EKEKE KKKKKKKKKKKKKKKKKKK KKEEE
      (K)(20)
75: KEKKK KKKKKKKKKKKKKKKKKKK KEEDE
      (K)(20)
76: EKKKK KKKKKKKKKKKKKKKKKKK EEEEA
      (R)(20)
133: GRRRS RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
134: RRRSR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
135: RRSRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
136: RSRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
137: SRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
138: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
139: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
140: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
141: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
142: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
143: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
144: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
145: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
146: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
147: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
148: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
149: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
150: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R,K)(20)
```

1

ABG26502 ck: 3399 len: 101 1 Abg26502 Novel human diagnostic protein

```
(R,K)(20)
59: EEEEG RRRRRRRRRRRRRRRRRRR RRRRR
      (R,K)(20)
60: EEGRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R,K)(20)
61: EGGRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R,K)(20)
62: EGRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R,K)(20)
63: GRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R,K)(20)
64: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R,K)(20)
65: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R,K)(20)
66: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R,K)(20)
67: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R,K)(20)
68: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R,K)(20)
69: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R,K)(20)
```

1

ABG26505 ck: 4704 len: 93 1 Abg26505 Novel human diagnostic protein

```
(R,K)(20)
60: KEEEG RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
61: EEGRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
62: EGGRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
63: EGRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
64: GRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
65: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R)(20)
66: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R,K)(20)
67: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R,K)(20)
68: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R,K)(20)
69: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R,K)(20)
```

1

ABG26506 ck: 3684 len: 85 ! Abg26506 Novel human diagnostic protein #26

(R,K){20}  
47: EKEKE KKKKKKKKKKKKKKKKKKK  
(R,K){20}  
48: KEKEK KKKKKKKKKKKKKKKKKKK  
(R,K){20}  
49: EKEER KKKKKKKKKKKKKKKKKKK  
(R,K){20}  
50: KEKKR KKKKKKKKKKKKKKKKKKK  
(R,K){20}  
51: EKRRR KKKKKKKKKKKKKKKKKKK  
(R,K){20}  
52: KRRRr KKKKKKKKKKKKKKKKKKK  
(K){20}  
53: RRRRr KKKKKKKKKKKKKKKKKKK  
(K){20}  
54: KRRRr KKKKKKKKKKKKKKKKKKK  
(K){20}  
55: RRRRr KKKKKKKKKKKKKKKKKKK  
(K){20}  
56: KRRRr KKKKKKKKKKKKKKKKKKK  
(K){20}  
57: RRRRr KKKKKKKKKKKKKKKKKKK  
(R,K){20}  
58: KRRRr KKKKKKKKKKKKKKKKKKK  
(R,K){20}  
59: KRRRr KKKKKKKKKKKKKKKKKKK  
(R,K){20}  
60: KRRRr KKKKKKKKKKKKKKKKKKK  
(R,K){20}  
61: KRRRr KKKKKKKKKKKKKKKKKKK  
(R,K){20}  
62: KRRRr KKKKKKKKKKKKKKKKKKK  
(R,K){20}

1

ABG26507 ck: 9838 len: 109 ! Abg26507 Novel human diagnostic protein #26

(R,K){20}  
60: EEEEE RRRRRRRRRRRRRRRRRRR  
(R){20}  
61: EEEEE RRRRRRRRRRRRRRRRRRR  
(R){20}  
62: EEEEE RRRRRRRRRRRRRRRRRRR  
(R){20}  
63: EEEEE RRRRRRRRRRRRRRRRRRR  
(R,K){20}  
64: EEEEE RRRRRRRRRRRRRRRRRRR  
(R,K){20}  
65: EEEEE RRRRRRRRRRRRRRRRRRR  
(R,K){20}

1

ABG26508 ck: 4488 len: 121 ! Abg26508 Novel human diagnostic protein

(R,K){20}  
56: DDEEE RRRKKKKKKKKKKKKKKKK  
(R,K){20}  
57: DEEEr RRRKKKKKKKKKKKKKKKK  
(R,K){20}  
58: DEEEr RRRKKKKKKKKKKKKKKKK  
(R,K){20}  
59: DEEEr RRRKKKKKKKKKKKKKKKK  
(R,K){20}  
60: EEEER KKKKKKKKKKKKKKKKKKK  
(R,K){20}  
61: RRRRr KKKKKKKKKKKKKKKKKKK  
(R,K){20}  
62: RRRRr KKKKKKKKKKKKKKKKKKK  
(R,K){20}  
63: RRRRr KKKKKKKKKKKKKKKKKKK  
(R,K){20}  
64: RRRRr KKKKKKKKKKKKKKKKKKK  
(R,K){20}  
65: RRRRr KKKKKKKKKKKKKKKKKKK  
(R,K){20}  
66: RRRRr KKKKKKKKKKKKKKKKKKK  
(R,K){20}  
67: RRRRr KKKKKKKKKKKKKKKKKKK  
(R,K){20}  
68: RRRRr KKKKKKKKKKKKKKKKKKK  
(R,K){20}  
69: RRRRr KKKKKKKKKKKKKKKKKKK  
(R,K){20}

1

ABG26510 ck: 3426 len: 74 ! Abg26510 Novel human diagnostic protein

(R,K){20}  
47: RRRRS RRRRRRRRRRRRRRRRRRR  
(R){20}  
48: RRRSR RRRRRRRRRRRRRRRRRRR  
(R){20}  
49: RRSRR RRRRRRRRRRRRRRRRRRR  
(R){20}  
50: RSRRR RRRRRRRRRRRRRRRRRRR  
(R){20}  
51: SRRRR RRRRRRRRRRRRRRRRRRR  
(R){20}

1

ABG26513 ck: 3117 len: 265 ! Abg26513 Novel human diagnostic protein

(R,K){20}  
199: EEEEE KKKKKKKKKKKKKKKKKKK  
(R,K){20}

222: KKEEE KKKKKKKKKKKKKKKKK KKEE  
(K){20}  
223: KEEKK KKKKKKKKKKKKKKKKK KKEE  
(R,K){20}  
224: EEEKK KKKKKKKKKKKKKKKKK KEEEE  
(R,K){20}  
225: EEKKK KKKKKKKKKKKKKKKKK EEEEE  
(R,K){20}

1  
ABG26514 ck: 9106 len: 218 ! Abg26514 Novel human diagnostic protein #26  
(R,K){20}  
(R){20}  
182: EEEGG RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
183: ERRGR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
184: RRGRR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
185: RGRRR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
186: GRRRR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
187: RRRRR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
188: RRRRR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
189: RRRRR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
190: RRRRR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
191: RRRRR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
192: RRRRR RRRRRRRRRRRRRRRRR RRRR  
(R,K){20}  
193: RRRRR RRRRRRRRRRRRRRRRR RRRR  
(R,K){20}

ABG26515 ck: 7288 len: 389 ! Abg26515 Novel human diagnostic protein #26  
(R,K){20}  
(R,K){20}  
259: KKEES KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
260: KEESR KKKKKKKKKKKKKKKKK KKKR  
(K){20}  
261: EESRK KKKKKKKKKKKKKKKKK KRRK  
(K){20}  
262: ESRRK KKKKKKKKKKKKKKKKK KRRK  
(R,K){20}  
263: SRKKK KKKKKKKKKKKKKKKKK KRRK  
(R,K){20}  
264: RKKKK KKKKKKKKKKKKKKKKK KRRK  
(R,K){20}

1  
265: KKKKK KKKKKKKKKKKKKKKKK KKKK  
(R,K){20}

ABG26516 ck: 2295 len: 91 ! Abg26516 Novel human diagnostic protein  
(R,K){20}  
(R,K){20}  
45: EEEEE RRRRRRRRRRRRRRRRR KRRK  
(R,K){20}  
46: EEEER RRRRRRRRRRRRRRRRR KRRR  
(R,K){20}  
47: EEEER RRRRRRRRRRRRRRRRR KRRK  
(R,K){20}  
48: EEEER RRRRRRRRRRRRRRRRR KRRK  
(R,K){20}  
49: EEEER RRRRRRRRRRRRRRRRR KRRK  
(R,K){20}  
50: EEEER RRRRRRRRRRRRRRRRR KRRK  
(R,K){20}  
51: EEEER RRRRRRRRRRRRRRRRR KRRK  
(R,K){20}  
52: EEEER RRRRRRRRRRRRRRRRR KRRK  
(R,K){20}  
53: EEEER RRRRRRRRRRRRRRRRR KRRK  
(R,K){20}  
54: EEEER RRRRRRRRRRRRRRRRR KRRK  
(R,K){20}

1  
ABG26518 ck: 8431 len: 761 ! Abg26518 Novel human diagnostic protein  
(R,K){20}  
(R){20}  
725: EEEEE RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
726: EEEER RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
727: EEEER RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
728: EEEER RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
729: EEEER RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
730: EEEER RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
731: EEEER RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
732: EEEER RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
733: EEEER RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
734: EEEER RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
735: EEEER RRRRRRRRRRRRRRRRR RRRR  
(R){20}



736: RRRR (R,K){20} RRRRRRRRRRRRRRRR KKKK  
737: RRRR (R,K){20} RRRRRRRRRRRRRRRR KKKK  
738: RRRR (R,K){20} RRRRRRRRRRRRRRRR KKKK  
739: RRRR (R,K){20} RRRRRRRRRRRRRRRR KKK  
740: RRRR (R,K){20} RRRRRRRRRRRRRRRR KN  
741: RRRR (R,K){20} RRRRRRRRRRRRRRRR N

ABG26520 ck: 2487 len: 99 1 Abg26520 Novel human diagnostic protein #26

1

35: EEEE (R,K){20} RRRRRRRRRRRRRRRR RRRR  
36: EEEE (R){20} RRRRRRRRRRRRRRRR RRRR  
37: EEEE (R){20} RRRRRRRRRRRRRRRR RRRR  
38: EEEE (R){20} RRRRRRRRRRRRRRRR RRRR  
39: EEEE (R){20} RRRRRRRRRRRRRRRR RRRR  
40: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
41: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
42: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
43: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
44: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
45: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
46: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
47: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
48: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
49: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
50: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
51: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR

52: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
53: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
54: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
55: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
56: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
57: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
58: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
59: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
60: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
61: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
62: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
63: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
64: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
65: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
66: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
67: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
68: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
69: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
70: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
71: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
72: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
73: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
74: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR  
75: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR

(R){20}  
76: RRRR RRRRRRRRRRRRRRRRRR RRKI  
(R){20}  
77: RRRR RRRRRRRRRRRRRRRRRR RKI  
(R){20}  
78: RRRR RRRRRRRRRRRRRRRRRR KI  
(R,K){20}  
79: RRRR RRRRRRRRRRRRRRRRRR I

1  
ABG26521 ck: 482 len: 367 ! Abg26521 Novel human diagnostic protein #26  
(R,K){20}  
317: EEEG KKKKKKKKKKKKKKKKK KKKR  
(K){20}  
318: EEEG KKKKKKKKKKKKKKKKK KKKR  
(K){20}  
319: EEGK KKKKKKKKKKKKKKKKK KRRR  
(K){20}  
320: EGKK KKKKKKKKKKKKKKKKK KRRK  
(K){20}  
321: GKKK KKKKKKKKKKKKKKKKK RRKK  
(R,K){20}  
322: KKKK KKKKKKKKKKKKKKKKK RRKK  
(R,K){20}  
323: KKKK KKKKKKKKKKKKKKKKK RRKK  
(R,K){20}  
324: KKKK KKKKKKKKKKKKKKKKK KKKK  
(R,K){20}  
325: KKKK KKKKKKKKKKKKKKKKK KKEQ  
(R,K){20}  
326: KKKK KKKKKKKKKKKKKKKKK KKEQ  
(R,K){20}  
327: KKKK KKKKKKKKKKKKKKKKK KEOL  
(R,K){20}  
328: KKKK KKKKKKKKKKKKKKKKK EQLE

1  
ABG26522 ck: 2060 len: 152 ! Abg26522 Novel human diagnostic protein #26  
(R,K){20}  
46: EEEE RRRRRRRRRRRRRRRRR RRKE  
(R){20}  
47: EEEER RRRRRRRRRRRRRRRRR RKKE  
(R){20}  
48: EEEER RRRRRRRRRRRRRRRRR KKEE  
(R,K){20}  
49: EEEER RRRRRRRRRRRRRRRRR KEGE  
(R,K){20}  
50: EERRR RRRRRRRRRRRRRRRRR EGEK  
(R,K){20}  
96: KKEG RRRRRRRRRRRRRRRRR RRKR

(R){20}  
97: KEEG RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
98: EEEG RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
99: EGER RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
100: GRRR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
101: RRRR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
102: RRRR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
103: RRRR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
104: RRRR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
105: RRRR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
106: RRRR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
107: RRRR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
108: RRRR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
109: RRRR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
110: RRRR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
111: RRRR RRRRRRRRRRRRRRRRR RRRR  
(R){20}  
112: RRRR RRRRRRRRRRRRRRRRR RRRR  
(R,K){20}  
113: RRRR RRRRRRRRRRRRRRRRR RRRR

1  
ABG26525 ck: 5433 len: 103 ! Abg26525 Novel human diagnostic protein  
(R,K){20}  
3: RD RRRRRRRRRRRRRRRRR EEEE  
(R,K){20}  
1  
ABG26526 ck: 1887 len: 115 ! Abg26526 Novel human diagnostic protein  
(R,K){20}  
48: EEEE RRRRRRRRRRRRRRRRR KRRK  
(R){20}  
49: EEEER RRRRRRRRRRRRRRRRR KRRK  
(R,K){20}  
50: EEEER RRRRRRRRRRRRRRRRR RRRR  
(R,K){20}

51: EERRR RRRRRRRRRRRRRRRRKKR EKKKS  
ABG26527 ck: 5586 len: 122 1 Abg26527 Novel human diagnostic protein #26  
(R,K){20}  
82: EEEEE KKKKKKKKKKKKKRRRRR GRRRM  
(R,K){20}  
1 ABG26528 ck: 2237 len: 215 1 Abg26528 Novel human diagnostic protein #26  
(R,K){20}  
87: EKEKE KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
88: KEKEK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
89: EKEKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
90: KEKKR KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
91: EKKRR KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
92: KKKRK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
93: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
94: RKKRR KKKKKKKKKKKKKKKKKKK KKEEE  
(K){20}  
95: KKKRK KKKKKKKKKKKKKKKKKKK KEEEE  
(K){20}  
96: KKKKK KKKKKKKKKKKKKKKKKKK EEEEE  
(K){20}  
167: GRRRG RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
168: RRRGR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
169: RRGRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
170: RGRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
171: GRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
172: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
173: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
174: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
175: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
176: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}

177: RRRRR RRRRRRRRRRRRRRRRRR MRKR  
(R){20}  
ABG26530 ck: 5729 len: 404 1 Abg26530 Novel human diagnostic protein  
(R,K){20}  
366: EEEEE RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
367: EEEER RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
368: EEEER RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
369: EEEER RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
370: ERRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
371: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
372: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
373: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
374: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
375: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
376: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
377: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
378: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
379: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
1 ABG26531 ck: 7434 len: 126 1 Abg26531 Novel human diagnostic protein  
(R,K){20}  
48: EEKDE RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
49: EKDER RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
50: KDEER RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
51: DEERR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
52: ERRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
53: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
54: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}

55: RRRR (R){20} RREE  
 56: RRRR (R){20} REEE  
 57: RRRR (R){20} EEEE  
 ABG26532 ck: 9650 len: 225 1 Abg26532 Novel human diagnostic protein #26  
 152: EEEE (R,K){20} RRRR  
 153: EEEK (R,K){20} RRRR  
 154: EEEK (R,K){20} RRRR  
 155: EKKK (R){20} RRRR  
 156: EKKK (R){20} RRRR  
 157: KKKR (R){20} RRRR  
 158: KKKR (R){20} RRRR  
 159: KRRR (R){20} RRRR  
 160: RRRR (R){20} RRRR  
 161: RRRR (R){20} RRRR  
 162: RRRR (R){20} RRRR  
 163: RRRR (R){20} RRRR  
 164: RRRR (R){20} RRRR  
 165: RRRR (R){20} RRRR  
 166: RRRR (R){20} RRRR  
 167: RRRR (R){20} RRRR  
 ABG26533 ck: 5900 len: 101 1 Abg26533 Novel human diagnostic protein #26  
 51: RRRR (R,K){20} RRRR  
 52: RRRR (R){20} RRRR

53: RRRR (R){20} RRRR  
 54: RRRR (R){20} RRRR  
 55: RRRR (R){20} RRRR  
 56: RRRR (R){20} RRRR  
 57: RRRR (R,K){20} RRRR  
 58: RRRR (R,K){20} RRRR  
 59: RRRR (R,K){20} RRRR  
 ABG26534 ck: 1919 len: 113 1 Abg26534 Novel human diagnostic protein  
 31: KKKK (R,K){20} KKKK  
 32: KKKK (R,K){20} KKKK  
 33: KKKK (R,K){20} KKKK  
 34: KKKK (R,K){20} KKKK  
 35: KKKK (R,K){20} KKKK  
 36: KKKK (R,K){20} KKKK  
 37: KKKK (R,K){20} KKKK  
 38: KKKK (R,K){20} KKKK  
 39: KKKK (R,K){20} KKKK  
 40: KKKK (R,K){20} KKKK  
 41: KKKK (R,K){20} KKKK  
 42: KKKK (R,K){20} KKKK  
 43: KKKK (R,K){20} KKKK  
 44: KKKK (R,K){20} KKKK  
 45: KKKK (R,K){20} KKKK  
 46: KKKK (R,K){20} KKKK

47: KKKK KKKKKKKKKKKKKKK KKKK  
(K){20}  
48: KKKK KKKKKKKKKKKKKKK KKKK  
(K){20}  
49: KKKK KKKKKKKKKKKKKKK KKKK  
(R,K){20}  
50: KKKK KKKKKKKKKKKKKKK KKKK  
(R,K){20}  
51: KKKK KKKKKKKKKKKKKKK KKKK  
(R,K){20}  
52: KKKK KKKKKKKKKKKKKKK KKKK  
(R,K){20}  
53: KKKK KKKKKKKKKKKKKKK KKKK  
(R,K){20}  
54: KKKK KKKKKKKKKKKKKKK KKKK  
(R,K){20}  
55: KKKK KKKKKKKKKKKKKKK KKKK  
(R,K){20}  
ABG26535 ck: 2214 len: 472 1 Abg26535 Novel human diagnostic protein #26  
1  
415: EEEE RRRRRRRRRRRRRRR RRRR  
(R){20}  
416: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
417: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
418: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
419: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
420: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
421: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
422: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
423: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
424: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
425: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
426: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
427: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}  
428: EEEER RRRRRRRRRRRRRRR RRRR  
(R){20}

429: RRRR RRRRRRRRRRRRRRR RRRR  
(R){20}  
430: RRRR RRRRRRRRRRRRRRR RRRR  
(R){20}  
431: RRRR RRRRRRRRRRRRRRR RRRR  
(R){20}  
432: RRRR RRRRRRRRRRRRRRR RRRR  
(R){20}  
433: RRRR RRRRRRRRRRRRRRR RRRR  
(R){20}  
434: RRRR RRRRRRRRRRRRRRR RRRR  
(R){20}  
435: RRRR RRRRRRRRRRRRRRR RRRR  
(R){20}  
436: RRRR RRRRRRRRRRRRRRR RRRR  
(R){20}  
437: RRRR RRRRRRRRRRRRRRR RRRR  
(R){20}  
438: RRRR RRRRRRRRRRRRRRR RRRR  
(R){20}  
ABG26537 ck: 2429 len: 573 1 Abg26537 Novel human diagnostic protein  
1  
405: GAGE RRRRRRRRRRRRRRR RREE  
(R,K){20}  
406: AGEER RRRRRRRRRRRRRRR RREE  
(R){20}  
407: EGER RRRRRRRRRRRRRRR EEEE  
(R){20}  
479: KEEE KKKKKKKKKKKKKKK KKKK  
(R,K){20}  
480: KEEE KKKKKKKKKKKKKKK KKKK  
(R,K){20}  
481: EEEK KKKKKKKKKKKKKKK KKKK  
(R,K){20}  
482: EEEK KKKKKKKKKKKKKKK KKKK  
(R){20}  
483: EEEK KKKKKKKKKKKKKKK KKKK  
(R){20}  
484: KKKK KKKKKKKKKKKKKKK KKKK  
(R){20}  
485: KKKK KKKKKKKKKKKKKKK KKKK  
(R){20}  
486: KKKK KKKKKKKKKKKKKKK KKKK  
(R){20}

1  
ABG26538 ck: 5732 len: 98 1 Abg26538 Novel human diagnostic protein  
(R,K){20}  
60: KEKE KRRRRRRRRRRRRRR KKKK  
(R,K){20}

```
(R,K){20}
61: EKETK RRRRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}
62: KETKR RRRRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}
63: ETKRK RRRRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}
64: TKRRR RRRRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}
65: KRRRR RRRRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}
66: RRRRR RRRRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}
67: RRRRK RRRRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}
68: RRRKK RRRRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}
69: RRRKK RRRRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}
70: RKKKK RRRRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}
71: KKKKK RRRRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}
72: KKKKR RRRRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}
73: KKKRK RRRRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}
74: KKRKK RRRRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}
75: KKKKK RRRRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}
76: RKKKK RRRRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}
77: KKKKK RRRRRRRRRRRRRRRRRRRR KKKKK
ABG26539 ck: 3545 len: 89 1 Abg26539 Novel human diagnostic protein #26
(R,K){20}
28: EEEEE RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
29: EEEER RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
30: EEEER RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
31: EEEER RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
32: EEEER RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
33: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
```

```
(R,K){20}
34: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
35: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
36: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
37: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
38: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
39: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
40: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
41: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
42: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
43: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
44: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
45: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
46: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
47: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
48: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
49: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
50: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
ABG26542 ck: 4605 len: 182 1 Abg26542 Novel human diagnostic protein
(R,K){20}
146: EEEEE RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
147: GEEER RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
148: EEEER RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
149: EEEER RRRRRRRRRRRRRRRRRRRR RRRRR
ABG26543 ck: 2998 len: 405 1 Abg26543 Novel human diagnostic protein
(R,K){20}
264: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
```

[illegible]

```

1
ABG26551  ck: 8976  len: 94  ! Abg26551 Novel human diagnostic protein

93: KKKKK (K){20} KKKKK
94: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
95: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
96: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
97: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
98: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

2: K RRRKKKKKKKKKKKKKKKKKK KKKKK
3: KR RKKKKKKKKKKKKKKKKKK KKKKK
4: KRR RKKKKKKKKKKKKKKKKKK KKKKK
5: KRRR KKKKKKKKKKKKKKKKKKK KKKKK
6: KRRRK KKKKKKKKKKKKKKKKKKK KKKKK
7: RRRKK KKKKKKKKKKKKKKKKKKK KKKKK
8: RRRKK KKKKKKKKKKKKKKKKKKK KKKKK
9: RKKKK KKKKKKKKKKKKKKKKKKK KKKKK
10: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
11: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
12: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
13: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
14: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
15: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
16: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
17: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK

```

18: KKKKK (R,K){20}  
19: KKKKK (R,K){20}  
20: KKKKK (R,K){20}  
21: KKKKK (R,K){20}

ABG26717 ck: 4415 len: 78 ! Abg26717 Novel human diagnostic protein #26

37: QMLSV (R,K){20}  
38: MLSVK (R,K){20}  
39: LSVKK (R,K){20}  
40: SVKKK (R,K){20}  
41: VKKKK (R,K){20}  
42: KKKKK (R,K){20}  
43: KKKKK (K){20}  
44: KKKKK (K){20}  
45: KKKKK (K){20}  
46: KKKKK (K){20}  
47: KKKKK (K){20}  
48: KKKKK (K){20}  
49: KKKKK (K){20}  
50: KKKKK (K){20}  
51: KKKKK (K){20}  
52: KKKKK (K){20}  
53: KKKKK (K){20}  
54: KKKKK (K){20}

ABG26718 ck: 9531 len: 141 ! Abg26718 Novel human diagnostic protein #26

1

80: EVARP (R,K){20}  
81: VARP (R,K){20}  
82: ARPRK (K){20}  
83: RPRKK (K){20}  
84: PRKKK (K){20}  
85: RKKKK (K){20}  
86: KKKKK (K){20}  
87: KKKKK (K){20}  
88: KKKKK (K){20}  
89: KKKKK (K){20}

ABG26719 ck: 72 len: 83 ! Abg26719 Novel human diagnostic protein

42: ETPSE (R,K){20}  
43: TPSEK (K){20}  
44: PSEKK (K){20}  
45: SEKKK (K){20}  
46: EKKKK (K){20}  
47: KKKKK (K){20}  
48: KKKKK (K){20}  
49: KKKKK (K){20}  
50: KKKKK (K){20}  
51: KKKKK (K){20}  
52: KKKKK (K){20}  
53: KKKKK (K){20}



1  
ABG26720 ck: 3781 len: 57 ! Abg26720 Novel human diagnostic protein #26

54: KKKK KKKKKKKKKKKKKKKKKKKKKK NPIFF

(R,K){20}

24: QMKST KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

25: MKSII KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

26: KSIIK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

27: SIKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

28: IKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

29: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

30: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

31: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

32: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

33: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

ABG26721 ck: 287 len: 95 ! Abg26721 Novel human diagnostic protein #26

1  
(R,K){20}

31: EKEKE KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

32: KEKEK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

33: EKEKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

34: KEKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

35: EKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

36: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

37: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

38: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

39: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

40: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

41: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

42: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

43: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

44: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

45: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

46: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

47: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

48: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

49: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

50: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

51: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

52: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

53: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

54: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

ABG26722 ck: 4831 len: 127 ! Abg26722 Novel human diagnostic protein

1

(R,K){20}

44: KKKKE KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

45: KKEEK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

46: KEEEK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

47: EEKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

48: EKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

49: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

87: EKEEE KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

88: KEEEK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

89: EDEKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

90: EEKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
91: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
92: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
93: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
94: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
95: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
96: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
97: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
98: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
99: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
100: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
101: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
102: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
103: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
104: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
105: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
106: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
107: KKKKK KKKKKKKKKKKKKKKKKKK K  
(K){20}  
108: KKKKK KKKKKKKKKKKKKKKKKKK  
ABG26723 ck: 7054 len: 98 ! Abg26723 Novel human diagnostic protein #26  
(R,K){120}  
42: EKEKE KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
43: KEKEK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
44: EKEKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
45: KEKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
46: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
50: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
51: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
52: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
53: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
54: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
55: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
56: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
57: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
58: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
59: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
60: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
61: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
62: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
63: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
64: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
65: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
66: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
67: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
68: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
69: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
70: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

1

ABG26724 ck: 9897 len: 295 ! Abg26724 Novel human diagnostic protein #26

(R,K){20}

259: KEEER KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

260: EEEER KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

261: EEEER KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

262: EEEER KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

263: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

264: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

265: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

266: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

267: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

268: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

269: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

270: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

271: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

272: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

273: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

274: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

275: KKKKK KKKKKKKKKKKKKKKKKKK K

(K){20}

276: KKKKK KKKKKKKKKKKKKKKKKKK K

1

ABG26725 ck: 909 len: 131 ! Abg26725 Novel human diagnostic protein #26

(R,K){20}

70: KKKKE KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

71: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

72: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

73: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

1

ABG26726 ck: 9300 len: 677 ! Abg26726 Novel human diagnostic protein

(R,K){20}

240: EKEKE KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

241: KEKER KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

242: EKEER KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

243: KEERK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

244: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

245: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

246: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

247: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

248: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

249: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

250: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

251: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

252: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

```

253: RKKKK (K){20}
254: KKKKK (K){20}
255: KKKKK (K){20}
256: KKKKK (K){20}
257: KKKKK (K){20}
258: KKKKK (K){20}
259: KKKKK (K){20}
260: KKKKK (K){20}
261: KKKKK (R,K){20}
262: KKKKK (R,K){20}
263: KKKKK (R,K){20}
264: KKKKK (R,K){20}
265: KKKKK (R,K){20}

ABG26727 ck: 813 len: 329 1 Abg26727 Novel human diagnostic protein #28
(R,K){20}
215: EKEKE KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
216: KEKEK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
217: EKEKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
218: KEKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
219: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
220: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
221: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
222: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
223: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
224: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
```

```

225: KKKKK (K){20}
226: KKKKK (K){20}
227: KKKKK (K){20}
228: KKKKK (K){20}
229: KKKKK (K){20}
230: KKKKK (K){20}

ABG26879 ck: 2669 len: 800 1 Abg26879 Novel human diagnostic protein
(R,K){20}
559: EREDE KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
560: REDEK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
561: DEKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
562: QEKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
563: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
564: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

ABG28885 ck: 8668 len: 137 1 Abg28885 Novel human diagnostic protein
(R,K){20}
35: EEEEG RRRRRRRRRRRRRRRRRRR RRRKK
(R,K){20}
36: EEEGR RRRRRRRRRRRRRRRRRRR RRRKK
(R){20}
37: EEEGR RRRRRRRRRRRRRRRRRRR RRRKK
(R){20}
38: EGRRR RRRRRRRRRRRRRRRRRRR RRRKK
(R){20}
39: GRRRR RRRRRRRRRRRRRRRRRRR RRRKK
(R){20}
40: RRRRR RRRRRRRRRRRRRRRRRRR RRRKK
(R,K){20}
41: RRRRR RRRRRRRRRRRRRRRRRRR RRRKK
(R,K){20}

AAG65985 ck: 8085 len: 154 1 Aag65985 B726P splice variant sequence.
(R,K){20}
114: TOLPO KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
```

115: QLRQK KKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
116: LROKK KKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
117: ROQKK KKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
118: OKKKK KKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
119: KKKKK KKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
120: KKKKK KKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
121: KKKKK KKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
122: KKKKK KKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
123: KKKKK KKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
124: KKKKK KKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
125: KKKKK KKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
126: KKKKK KKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
127: KKKKK KKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
128: KKKKK KKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

ABb27893 ck: 5383 len: 86 1 Abb27893 Human peptide #544 encoded by brea

(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}  
(R,K){20}

1

24: RKKKK KKKKKKKKKKKKKKKKKK RRRRR  
(R,K){20}  
25: KKKKK KKKKKKKKKKKKKKKKKK RRRRR  
(R,K){20}  
26: KKKKK KKKKKKKKKKKKKKKKKK RRRRR  
(R,K){20}  
27: KKKKK KKKKKKKKKKKKKKKKKK RRRRR  
(R,K){20}  
28: KKKKK KKKKKKKKKKKKKKKKKK RRRRR  
(R,K){20}  
29: KKKKK KKKKKKKKKKKKKKKKKK RRRRR  
(R,K){20}  
30: KKKKK KKKKKKKKKKKKKKKKKK RRRRR  
(R,K){20}  
31: RKKKK KKKKKKKKKKKKKKKKKK RRRRR  
(R,K){20}  
32: KKKKK KKKKKKKKKKKKKKKKKK RRRRR  
(R,K){20}  
33: KKKKK KKKKKKKKKKKKKKKKKK RRRRR  
(R,K){20}  
34: KKKKK KKKKKKKKKKKKKKKKKK RRRRR  
(R,K){20}  
35: KKKKK KKKKKKKKKKKKKKKKKK RRRRR  
(R,K){20}  
36: KKKKK KKKKKKKKKKKKKKKKKK RRRRR  
(R,K){20}  
37: RKKKK RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
38: KKKKK RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
39: KKKKK RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
40: KKKKK RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
41: KRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
42: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
43: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
44: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
45: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R){20}

ABb28750 ck: 1334 len: 86 1 Abb28750 Peptide #1401 encoded by brea

(R,K){20}  
(R,K){20}  
(R,K){20}

```

(R,K){20}
45: KERRR RRRRRRRRRRRRRRRRRR RRKK
(R,K){20}
46: ERRR RRRRRRRRRRRRRRRRRR RRKK
(R,K){20}
47: RRRR RRRRRRRRRRRRRRRR RKKK
(R,K){20}
48: RRRR RRRRRRRRRRRRRRRR KKKK
(R,K){20}
49: RRRR RRRRRRRRRRRRRRRR KKKK
(R,K){20}
50: RRRR RRRRRRRRRRRRRRRR KKKE
(R,K){20}
51: RRRR RRRRRRRRRRRRRRRR KKKE
(R,K){20}
52: RRRR RRRRRRRRRRRRRRRR KKEE
(R,K){20}
53: RRRR RRRRRRRRRRRRRRRR KEEE
(R,K){20}
54: RRRR RRRRRRRRRRRRRRRR EEEE

ABb29928 ck: 3607 len: 88 ! Abb29928 Peptide #2579 encoded by breast
(R,K){20}
39: ERKRE KKKKKKKKKKKKKKKK KKKK
(K){20}
40: RKRER KKKKKKKKKKKKKKKK KKKK
(K){20}
41: KRERK KKKKKKKKKKKKKKKK KKKK
(K){20}
42: REKKR KKKKKKKKKKKKKKKK KKKK
(K){20}
43: EKKKR KKKKKKKKKKKKKKKK KKKK
(K){20}
44: KKKRK KKKKKKKKKKKKKKKK KKKK
(K){20}
45: KKKRK KKKKKKKKKKKKKKKK KKKK
(K){20}
46: KKKRK KKKKKKKKKKKKKKKK KKKK
(K){20}
47: KKKRK KKKKKKKKKKKKKKKK KKKK
(K){20}
48: KKKRK KKKKKKKKKKKKKKKK KKKK
(K){20}
49: KKKRK KKKKKKKKKKKKKKKK KKKK
(K){20}
50: KKKRK KKKKKKKKKKKKKKKK KKKK
(K){20}
51: KKKRK KKKKKKKKKKKKKKKK KKKK

```

52: KKKK (K){20} KKKK  
53: KKKK (K){20} KKKK  
54: KKKK (K){20} KKKK  
55: KKKK (K){20} KKKK  
56: KKKK (K){20} KKKK  
57: KKKK (K){20} KKKK  
58: KKKK (K){20} KKKK  
59: KKKK (K){20} KKKK  
60: KKKK (K){20} KKKK  
61: KKKK (K){20} KKKK  
62: KKKK (K){20} KKKK  
63: KKKK (K){20} KKKK  
64: KKKK (K){20} KKKK  
65: KKKK (K){20} KKKK  
66: KKKK (K){20} KKKK  
67: KKKK (K){20} KKKK  
68: KKKK (K){20} K  
69: KKKK (K){20} KKKK  
AB030512 ck: 3937 len: 85 ! Ab030512 Peptide #3163 encoded by breast c  
(R,K){20}  
1: KKKK (K){20} KKKK  
2: K (K){20} KKKK  
3: K (K){20} KKKK  
4: K (K){20} KKKK  
5: K (K){20} KKKK

6: KKKK (K){20} KKKK  
7: KKKK (K){20} KKKK  
8: KKKK (K){20} KKKK  
9: KKKK (K){20} KKKK  
10: KKKK (K){20} KKKK  
11: KKKK (K){20} KKKK  
AB032308 ck: 1560 len: 88 ! Ab032308 Peptide #4959 encoded by breas  
(R,K){20}  
43: RRRG (R){20} RRRR  
44: ERGR (R){20} RRRR  
45: RGRG (R){20} RRRR  
46: RGRG (R){20} RRRR  
47: GRGR (R){20} RRRR  
48: RRRR (R){20} RRRR  
49: RRRR (R){20} RRRR  
50: RRRR (R){20} RRRR  
51: RRRR (R){20} RRRR  
52: RRRR (R){20} RRRR  
53: RRRR (R){20} RRRR  
54: RRRR (R){20} RRRR  
55: RRRR (R){20} RRRR  
56: RRRR (R){20} RRRR  
57: RRRR (R){20} RRRR  
58: RRRR (R){20} RRRR  
59: RRRR (R){20} RRRR

60: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRTN  
61: RRRR (R){20} RRRRRRRRRRRRRRRRR RNTNN  
62: RRRR (R){20} RRRRRRRRRRRRRRRRR NTNNE  
ABB33064 ck: 5383 len: 86 ! Abb33064 Peptide #570 encoded by human foet  
15: RRRG (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
16: RRRG (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
17: RRRG (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
18: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
19: GRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
20: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
21: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
22: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
23: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
24: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
25: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
26: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
27: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
28: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
29: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
30: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
31: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
32: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
33: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
34: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR

35: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
36: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
37: RRRR (R){20} RRRRRRRRRRRRRRRRR KRRR  
38: RRRR (R){20} RRRRRRRRRRRRRRRRR KRRR  
39: RRRR (R){20} RRRRRRRRRRRRRRRRR KRRR  
40: RRRR (R){20} RRRRRRRRRRRRRRRRR KRRR  
41: RRRR (R){20} RRRRRRRRRRRRRRRRR KRRR  
42: RRRR (R){20} RRRRRRRRRRRRRRRRR KRRR  
43: RRRR (R){20} RRRRRRRRRRRRRRRRR KRRR  
44: RRRR (R){20} RRRRRRRRRRRRRRRRR KRRR  
45: RRRR (R){20} RRRRRRRRRRRRRRRRR KRRR  
1  
ABB33937 ck: 1334 len: 86 ! Abb33937 Peptide #1443 encoded by human  
57: EEEG (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
58: EEEG (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
59: EEEG (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
60: EEEG (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
61: GRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
62: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
63: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
64: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
65: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
66: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
67: RRRR (R,K){20} RRRRRRRRRRRRRRRRR KRRR  
ABB34024 ck: 9082 len: 167 ! Abb34024 Peptide #1530 encoded by human



1

(R,K){20}  
(R,K){20}  
33: EEGRG RRRRRRRRRRRRRRRRRR RRGCG  
(R,K){20}  
34: EGRGR RRRRRRRRRRRRRRRRRR RGGGR  
(R,K){20}  
35: GRGRR RRRRRRRRRRRRRRRRRR GGGRR

1

ABB34533 ck: 2276 len: 89 1 Abb34533 Peptide #2039 encoded by human fo  
(R,K){20}  
(R,K){20}  
23: EEEEE KKKKKKKKKKKKKKKKK EEEK  
(K){20}  
46: KKEEE KKKKKKKKKKKKKKKKK KRKEE  
(K){20}  
47: KEEER KKKKKKKKKKKKKKKKK KKEEE  
(R,K){20}  
48: EEEKK KKKKKKKKKKKKKKKKK KEEBE  
(R,K){20}  
49: EKKKK KKKKKKKKKKKKKKKKK EEEEE

1

ABB34819 ck: 1939 len: 130 1 Abb34819 Peptide #2325 encoded by human fo  
(R,K){20}  
(R,K){20}  
42: EGRKE RRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
43: GRKER RRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
44: RKERR RRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
45: KERRR RRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
46: EERRR RRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
47: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
48: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
49: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
50: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
51: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
52: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
53: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
54: RRRRR RRRRRRRRRRRRRRRRRR RRRRR

1

ABB35110 ck: 3607 len: 88 1 Abb35110 Peptide #2616 encoded by human  
(R,K){20}  
(K){20}  
39: ERRRE KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
40: RKREK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
41: KREKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
42: REKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
43: EKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
44: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
45: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(R){20}  
46: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
47: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(R){20}  
48: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
49: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
50: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
51: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
52: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
53: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
54: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
55: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
56: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
57: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
58: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
59: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
60: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
61: KKKKK KKKKKKKKKKKKKKKKK KKKKK

```
(K)[20]
62: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
63: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
64: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
65: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K)[20]
66: KKKKK KKKKKKKKKKKKKKKKKKK KKK
(K)[20]
67: KKKKK KKKKKKKKKKKKKKKKKKK KK
(K)[20]
68: KKKKK KKKKKKKKKKKKKKKKKKK K
(K)[20]
69: KKKKK KKKKKKKKKKKKKKKKKKK

ABD35676  3937  len: 85  1  Abd35676 Peptide #3182 encoded by human foe
(R,K)[20]
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
2: K KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK EEEEX
(R,K)[20]
ABD36406  2686  len: 71  1  Abd36406 Peptide #3912 encoded by human foe
(R,K)[20]
20: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
21: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
22: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
```

```
(K)[20]
23: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K)[20]
24: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
25: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
26: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
27: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
28: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
29: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
30: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
31: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
32: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
33: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
34: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
35: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
36: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
37: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
38: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
39: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
40: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
41: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
42: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
43: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
44: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
45: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
46: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKSA
```

(K){20}  
47: KKKR KKKKKKKKKKKKKKK KSAH  
(K){20}  
48: KKKR KKKKKKKKKKKKKKK KSAH  
(K){20}  
49: KKKR KKKKKKKKKKKKKKK SAH  
ABB37567 ck: 1560 len: 88 1 Abb37567 Peptide #5073 encoded by human fo  
(R,K){20}  
43: RRRG RRRRRRRRRRRRRRRR RRRR  
(R){20}  
44: ERGR RRRRRRRRRRRRRRRR RRRR  
(R){20}  
45: RGRR RRRRRRRRRRRRRRRR RRRR  
(R){20}  
46: RGRR RRRRRRRRRRRRRRRR RRRR  
(R){20}  
47: GRRR RRRRRRRRRRRRRRRR RRRR  
(R){20}  
48: RRRR RRRRRRRRRRRRRRRR RRRR  
(R){20}  
49: RRRR RRRRRRRRRRRRRRRR RRRR  
(R){20}  
50: RRRR RRRRRRRRRRRRRRRR RRRR  
(R){20}  
51: RRRR RRRRRRRRRRRRRRRR RRRR  
(R){20}  
52: RRRR RRRRRRRRRRRRRRRR RRRR  
(R){20}  
53: RRRR RRRRRRRRRRRRRRRR RRRR  
(R){20}  
54: RRRR RRRRRRRRRRRRRRRR RRRR  
(R){20}  
55: RRRR RRRRRRRRRRRRRRRR RRRR  
(R){20}  
56: RRRR RRRRRRRRRRRRRRRR RRRR  
(R){20}  
57: RRRR RRRRRRRRRRRRRRRR RRRR  
(R){20}  
58: RRRR RRRRRRRRRRRRRRRR RRRR  
(R){20}  
59: RRRR RRRRRRRRRRRRRRRR RRRR  
(R){20}  
60: RRRR RRRRRRRRRRRRRRRR RRRR  
(R){20}  
61: RRRR RRRRRRRRRRRRRRRR RRRR  
(R){20}  
62: RRRR RRRRRRRRRRRRRRRR RRRR

1  
ABB37780 ck: 2324 len: 36 1 Abb37780 Peptide #5286 encoded by human  
(R,K){20}  
16: KRRR KRRRRRRRRRRRRRRR R  
(R,K){20}  
17: ERRT KRRRRRRRRRRRRRRR

ABB40272 ck: 8343 len: 66 1 Abb40272 Peptide #7778 encoded by human

(R,K){20}  
6: EERE KKKKKKKKKKKKKKKKK KKKK  
(R,K){20}  
7: TERE KKKKKKKKKKKKKKKKK KKKK  
(R,K){20}  
8: EERE KKKKKKKKKKKKKKKKK KKKK  
(R,K){20}  
9: REKK KKKKKKKKKKKKKKKKK KKKK

ABB42642 ck: 2394 len: 57 1 Abb42642 Peptide #10148 encoded by huma

(R,K){20}  
20: EEEG RRRRRRRRRRRRRRRR RRRR  
(R){20}  
21: EEEG RRRRRRRRRRRRRRRR RRRR  
(R){20}  
22: EEEG RRRRRRRRRRRRRRRR RRRR  
(R){20}  
23: EGRR RRRRRRRRRRRRRRRR RRRR  
(R){20}  
24: GRRR RRRRRRRRRRRRRRRR RRRR  
(R){20}  
25: RRRR RRRRRRRRRRRRRRRR RRRR

ABB43181 ck: 4228 len: 24 1 Abb43181 Peptide #10687 encoded by huma

(R,K){20}  
1: RRRRRRRRRRRRRRRRRR RRRR  
(R,K){20}  
2: RRRRRRRRRRRRRRRRRR RRRR  
(R,K){20}  
3: RR RRRRRRRRRRRRRRRR RRRR  
(R,K){20}  
4: RRR RRRRRRRRRRRRRRRR RRRR

ABB44317 ck: 4895 len: 51 1 Abb44317 Peptide #11823 encoded by huma

(R,K){20}  
18: LKPM KRRRRRRRRRRRRRRR RRRR  
(R,K){20}  
19: FKPM KRRRRRRRRRRRRRRR RRRR

(R,K){20}  
20: KPMRK RRRKKRRRRKKKKRRKK LTTT  
ABD17165 ck: 8887 len: 42 ! Abd17165 Human nervous system related polyA  
(R,K){20}  
(K){20}  
21: FTTE KKKKKKKKKKKKKKKK KX  
(K){20}  
22: FTTEK KKKKKKKKKKKKKKKK X  
ABD18534 ck: 5383 len: 86 ! Abd18534 Protein #533 encoded by probe for  
(R,K){20}  
(R,K){20}  
15: RRRRG RRRRRKKKKKKKKKKKK KRRR  
(R,K){20}  
16: RRRGR RRRKKKKKKKKKKKKKK KRRR  
(R,K){20}  
17: RRGRR RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
18: RGRRR RRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
19: GRRRR RKKKKKKKKKKKKKKKK RRRR  
(R,K){20}  
20: RRRRR KKKKKKKKKKKKKKKKK RRRR  
(R,K){20}  
21: RRRRK KKKKKKKKKKKKKKKKK RRRR  
(R,K){20}  
22: RRRKK KKKKKKKKKKKKKKKKK RRRR  
(R,K){20}  
23: RRRKK KKKKKKKKKKKKKKKKK RRRR  
(R,K){20}  
24: RRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
25: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
26: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
27: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
28: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
29: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
30: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
31: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
32: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}

1  
33: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
34: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
35: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
36: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
37: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
38: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
39: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
40: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
41: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
42: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
43: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
44: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
45: KRRKK KRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
ABD19373 ck: 1334 len: 86 ! Abd19373 Protein #1372 encoded by probe  
(R,K){20}  
(R,K){20}  
57: EEEGG RRRKKKKKKKKKKKKKK KRRR  
(R,K){20}  
58: EEEGR RRRKKKKKKKKKKKKKK KRRR  
(R,K){20}  
59: EEEGR RRRKKKKKKKKKKKKKK KRRR  
(R,K){20}  
60: EEEGR RRRKKKKKKKKKKKKKK KRRR  
(R,K){20}  
61: GRRRK KRRKKKKKKKKKKKKKK KRRR  
(R,K){20}  
62: RRRKK KRRKKKKKKKKKKKKKK KRRR  
(R,K){20}  
63: RRRKK KRRKKKKKKKKKKKKKK KRRR  
(R,K){20}  
64: RRRKK KRRKKKKKKKKKKKKKK KRRR  
(R,K){20}  
65: KRRKK KRRKKKKKKKKKKKKKK KRRR  
(R,K){20}  
66: KRRKK KRRKKKKKKKKKKKKKK KRRR  
(R,K){20}

67: KKKRR KKKKKKKKKKKKKKKKK

ABBI9467 ck: 9082 len: 167 ! Abbi9467 Protein #1466 encoded by probe for

(R,K){20}

(R,K){20}

33: EEGRG RRRRRRRRRRRRRRRRRK RRGGR

(R,K){20}

34: EGRGR RRRRRRRRRRRRRRRRRK RGGGR

(R,K){20}

35: GGRGR RRRRRRRRRRRRRRRRRK GGGRR

ABBI9943 ck: 2276 len: 89 ! Abbi9943 Protein #1942 encoded by probe for

(R,K){20}

(R,K){20}

23: EEEED KKKKKKKKKKKKKKKKKK EEEKK

(K){20}

46: KKEED KKKKKKKKKKKKKKKKKK KKEED

(K){20}

47: KEEED KKKKKKKKKKKKKKKKKK KKEED

(R,K){20}

48: EEEKK KKKKKKKKKKKKKKKKKK KEEED

(R,K){20}

49: EEKKR KKKKKKKKKKKKKKKKKK EEEED

ABBI2038 ck: 1939 len: 130 ! Abbi2038 Protein #2237 encoded by probe for

(R,K){20}

(R,K){20}

42: EGRRR RRRRRRRRRRRRRRRRRK RRRRR

(R,K){20}

43: GRRRR RRRRRRRRRRRRRRRRRK RRRRR

(R,K){20}

44: EKKRR RRRRRRRRRRRRRRRRRK RRRRR

(R,K){20}

45: KERRR RRRRRRRRRRRRRRRRRK RRRRR

(R,K){20}

46: ERRRR RRRRRRRRRRRRRRRRRK RRRRR

(R,K){20}

47: RRRRR RRRRRRRRRRRRRRRRRK RRRRR

(R,K){20}

48: RRRRR RRRRRRRRRRRRRRRRRK RRRRR

(R,K){20}

49: RRRRR RRRRRRRRRRRRRRRRRK RRRRR

(R,K){20}

50: RRRRR RRRRRRRRRRRRRRRRRK RRRRR

(R,K){20}

51: RRRRR RRRRRRRRRRRRRRRRRK RRRRR

(R,K){20}

52: RRRRR RRRRRRRRRRRRRRRRRK RRRRR

(R,K){20}

53: RRRRR RRRRRRRRRRRRRRRRRK RRRRR

(R,K){20}

54: RRRRR RRRRRRRRRRRRRRRRRK EEEED

ABBI20531 ck: 3607 len: 88 ! Abbi20531 Protein #2530 encoded by probe for

(R,K){20}

(K){20}

39: EKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

40: KKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

41: KKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

42: RKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

43: EKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

44: KKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

45: KKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

46: KKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

47: KKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

48: KKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

49: KKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

50: KKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

51: KKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

52: KKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

53: KKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

54: KKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

55: KKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

56: KKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

57: KKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

58: KKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

59: KKKRR KKKKKKKKKKKKKKKKKK KKKKK

(K){20}

60: KKKRR KKKKKKKKKKKKKKKKKK KKKKK

61: KKKKK (K){20} KKKKK  
 62: KKKKK (K){20} KKKKK  
 63: KKKKK (K){20} KKKKK  
 64: KKKKK (K){20} KKKKK  
 65: KKKKK (K){20} KKKKK  
 66: KKKKK (K){20} KKKKK  
 67: KKKKK (K){20} KKKKK  
 68: KKKKK (K){20} KKKKK  
 69: KKKKK (K){20} KKKKK

ABB21105 ck: 3937 len: 85 1 Abb21105 Protein #3104 encoded by probe for

1: (R,K){20} KKKKK  
 2: K (K){20} KKKKK  
 3: KK (K){20} KKKKK  
 4: KKK (K){20} KKKKK  
 5: KKKK (K){20} KKKKK  
 6: KKKKK (K){20} KKKKK  
 7: KKKKK (K){20} KKKKK  
 8: KKKKK (K){20} KKKKK  
 9: KKKKK (K){20} KKKKK  
 10: KKKKK (K){20} KKKKK  
 11: KKKKK (K){20} KKKKK  
 ABB21763 ck: 2686 len: 71 1 Abb21763 Protein #3762 encoded by probe for  
 (R,K){20} KKKKK  
 20: KKKKK (K){20} KKKKK

21: KKKKK (K){20} KKKKK  
 22: KKKKK (K){20} KKKKK  
 23: KKKKK (K){20} KKKKK  
 24: KKKKK (K){20} KKKKK  
 25: KKKKK (R,K){20} KKKKK  
 26: KKKKK (R,K){20} KKKKK  
 27: KKKKK (R,K){20} KKKKK  
 28: KKKKK (R,K){20} KKKKK  
 29: KKKKK (R,K){20} KKKKK  
 30: KKKKK (R,K){20} KKKKK  
 31: KKKKK (R,K){20} KKKKK  
 32: KKKKK (R,K){20} KKKKK  
 33: KKKKK (R,K){20} KKKKK  
 34: KKKKK (R,K){20} KKKKK  
 35: KKKKK (R,K){20} KKKKK  
 36: KKKKK (R,K){20} KKKKK  
 37: KKKKK (R,K){20} KKKKK  
 38: KKKKK (R,K){20} KKKKK  
 39: KKKKK (R,K){20} KKKKK  
 40: KKKKK (R,K){20} KKKKK  
 41: KKKKK (R,K){20} KKKKK  
 42: KKKKK (R,K){20} KKKKK  
 43: KKKKK (R,K){20} KKKKK  
 44: KKKKK (R,K){20} KKKKK  
 45: KKKKK (R,K){20} KKKKK

(R,K){20}  
46: KKKKK KKKKKKKKKKKKKKKKKKK KKKSA  
(K){20}  
47: KKKKK KKKKKKKKKKKKKKKKKKK KKSAA  
(K){20}  
48: KKKKK KKKKKKKKKKKKKKKKKKK KSAH  
(K){20}  
49: KKKKK KKKKKKKKKKKKKKKKKKK SAH  
AB22862 ck: 1560 len: 88 ! Ab22862 Protein #4861 encoded by probe for  
(R,K){20}  
43: RRRGG RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
44: ERGGG RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
45: RGGRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
46: RGGRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
47: GRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
48: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
49: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
50: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
51: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
52: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
53: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
54: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
55: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
56: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
57: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
58: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
59: RRRRR RRRRRRRRRRRRRRRRRRR RRRRT  
(R){20}  
60: RRRRR RRRRRRRRRRRRRRRRRRR RRRTN  
(R){20}  
61: RRRRR RRRRRRRRRRRRRRRRRRR RRTNN

(R){20}  
62: RRRRR RRRRRRRRRRRRRRRRRRR NTNNE  
AB23064 ck: 2324 len: 36 ! Ab23064 Protein #5063 encoded by probe  
(R,K){20}  
16: KKKKT KKKRRRRRRRRRRRRRRRR R  
(R,K){20}  
17: ERKTK KKKRRRRRRRRRRRRRRRR  
AB24685 ck: 8343 len: 66 ! Ab24685 Protein #6684 encoded by probe  
(R,K){20}  
6: ETERE KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
7: TEREK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
8: EREKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
9: REKKK KKKKKKKKKKKKKKKKKKK NKKKK  
AB25988 ck: 2394 len: 57 ! Ab25988 Protein #7987 encoded by probe  
(R,K){20}  
20: EEEEG RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
21: EEEGR RRRRRRRRRRRRRRRRRRR RRRRG  
(R){20}  
22: EEEGR RRRRRRRRRRRRRRRRRRR RRRGR  
(R){20}  
23: EGGRR RRRRRRRRRRRRRRRRRRR RRRGR  
(R){20}  
24: GRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
25: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
AB27176 ck: 4895 len: 51 ! Ab27176 Protein #9175 encoded by probe  
(R,K){20}  
18: LKKPM KKKRRRRRRRRRRRRRRRR KLLTT  
(R,K){20}  
19: FKPMR KKKRRRRRRRRRRRRRRRR KLLTT  
(R,K){20}  
20: KPMRK KKKRRRRRRRRRRRRRRRR LTTT  
(R,K){20}  
AB210296 ck: 3983 len: 292 ! Ab210296 Human cDNA SEQ ID NO: 604. 1/2  
(R,K){20}  
273: QVFAP KKKKKKKKKKKKKKKKKKK

```
1 ABB10485   ck: 7611   len: 315   ! Abb10485 Human cDNA SEQ ID NO: 793. 1/2002
      (R,K){20}
273: QVEAP  RRRRRRRRRRRRRRRRRRRRRR KGGRS
      (K){20}
274: VFAPR  KKKKKKKKKKKKKKKKKKKKK GGRSR

1 AAU21948   ck: 444   len: 66   ! AAU21948 Human cardiovascular system antigen
      (R,K){20}
35: SMTPS  KKKKKKKKKKKKKKKKKKKKK KXGKK
      (K){20}
36: MTFBK  KKKKKKKKKKKKKKKKKKKKK XGKKK

1 AAU22148   ck: 4665  len: 34   ! AAU22148 Human cardiovascular system antigen
      (R,K){20}
10: PELIL  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
11: ELLLK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
12: LLLKK  KKKKKKKKKKKKKKKKKKKKK KKK
      (K){20}
13: LLKKK  KKKKKKKKKKKKKKKKKKKKK KK
      (K){20}
14: LKKKK  KKKKKKKKKKKKKKKKKKKKK K
      (K){20}
15: KKKKK  KKKKKKKKKKKKKKKKKKKKK

1 AAU22186   ck: 269   len: 76   ! AAU22186 Human cardiovascular system antigen
      (R,K){20}
37: TPSRA  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
38: PSRAK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
39: SRAKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
40: RAKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
41: AKKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
42: KKKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
43: KKKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
44: KKKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK

1 AAU22374   ck: 8278  len: 53   ! AAU22374 Human cardiovascular system antigen
      (R,K){20}

1
30: NCGIL  KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
31: CGILK  KKKKKKKKKKKKKKKKKKKKK KKK
      (K){20}
32: GILKK  KKKKKKKKKKKKKKKKKKKKK KK
      (K){20}
33: ILKKK  KKKKKKKKKKKKKKKKKKKKK K
      (K){20}
34: LKKKK  KKKKKKKKKKKKKKKKKKKKK

1 AAU23799   ck: 6158  len: 272   ! AAU23799 Novel human enzyme polypeptide
      (R,K){20}
238: SFANA  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
239: PANAK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
240: ANAKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
241: NAKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
242: AKKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK

1 AAU27944   ck: 1121  len: 69   ! AAU27944 Human config polypeptide sequen
      (R,K){20}
40: VPPLT  RKKKKKKKKKKKKKKKKKKKK KKKKK
      (R,K){20}
41: PPLTR  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
42: PLTRK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
43: LTRKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
44: TRKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
45: RKKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
46: KKKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (R,K){20}
47: KKKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK

1 AAU31467   ck: 4264  len: 657   ! AAU31467 Novel human secreted protein #1
      (R,K){20}
19: RRRRP  RRRRRRRRRRRRRRRRRRRRR RRRRL
      (R){20}
20: RRRRP  RRRRRRRRRRRRRRRRRRRRR RRRRL
      (R){20}
```



21: RPRRR RRRRRRRRRRRRRRRRRR RRLGL  
(R){20}  
22: RPRRR RRRRRRRRRRRRRRRRRR RLGLE  
(R){20}  
23: PRRRR RRRRRRRRRRRRRRRRRR LGLER

AAU3348 ck: 8085 len: 154 1 Aau3348 Human breast cancer protein encode

1

(R,K){20}  
114: TQLHQ KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

(K){20}  
115: QLRQK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
116: LROKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
117: RQKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
118: QKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
119: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
120: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
121: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
122: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
123: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
124: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
125: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
126: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
127: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
128: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

AAE09664 ck: 1663 len: 87 1 Aae09664 Human pancreatic related protein H

1

(R,K){20}  
(K){20}

36: KWSXX KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
37: WSSXX KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
38: SSXXK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
39: SXKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
40: XXXKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
41: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
42: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
43: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
50: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
51: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
52: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
53: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
54: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
55: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
56: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
57: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
58: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
59: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
60: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
61: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
62: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
63: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

```
AAW95365  ck: 5626  len: 139  ! Aaw95365 Human reproductive system related
1
(R,K){20}
111: IHLNL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
112: HLNLK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
113: LNLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
114: NLKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

AAM96607  ck: 4751  len: 80  ! Aam96607 Human reproductive system related
1
(R,K){20}
61: KXXFD KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

AAU18162  ck: 7907  len: 39  ! Aau18162 Novel human DNA-binding protein #9
1
(R,K){20}
9: YFEDL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
10: FEDLK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
11: EDLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
12: DLKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
13: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

AAU18167  ck: 9194  len: 87  ! Aau18167 Novel human DNA-binding protein #1
1
(R,K){20}
52: KIILL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
53: IILKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
54: ILKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
55: LLKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
56: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

AAU18168  ck: 8659  len: 104  ! Aau18168 Novel human DNA-binding protein #1
1
(R,K){20}
75: PLGGQ KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

(K){20}
76: LGGQK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
77: GGQKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
78: GQKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
79: QKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
80: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

AAU18171  ck: 9398  len: 48  ! Aau18171 Novel human DNA-binding protein
1
(R,K){20}
2: Q KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
3: QK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
4: QKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

AAU18177  ck: 8278  len: 53  ! Aau18177 Novel human DNA-binding protein
1
(R,K){20}
30: NCGIL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
31: CGILK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
32: GILKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
33: ILKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
34: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

AAU18178  ck: 444  len: 66  ! Aau18178 Novel human DNA-binding protein
1
(R,K){20}
35: SMTFS KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
36: MTFSS KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

AAU18179  ck: 5503  len: 50  ! Aau18179 Novel human DNA-binding protein
1
(R,K){20}
30: ICCLL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
31: ICCLK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

AAU18184  ck: 5691  len: 108  ! Aau18184 Novel human DNA-binding protein
1
(R,K){20}
78: VRPCL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
```

79: RPCLK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
80: PCLKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
81: CLKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
82: LKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
83: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
84: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
85: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
86: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
87: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
88: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
89: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
1  
AAU18192 ck: 6029 len: 63 1 Aau18192 Novel human DNA-binding protein #3  
(R,K){20}  
(K){20}  
40: RLTL KKKKKKKKKKKKKKKKKKKKK ISWG  
1  
AAU18200 ck: 7170 len: 63 1 Aau18200 Novel human DNA-binding protein #4  
(R,K){20}  
(K){20}  
37: TPSRA KKKKKKKKKKKKKKKKKKKKK KKKKK  
38: PSRAK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
39: SRAKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
40: RAKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
41: AKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
42: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
43: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
44: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
AAU18204 ck: 6110 len: 61 1 Aau18204 Novel human DNA-binding protein #5  
(R,K){20}

1  
28: RPTRP KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
29: PTRPK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
30: TRPKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
31: RPKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
32: PKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
33: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
34: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
35: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
36: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
37: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
AAU18205 ck: 5764 len: 74 1 Aau18205 Novel human DNA-binding protei  
(R,K){20}  
(K){20}  
40: EPLSA KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
41: FLSAK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
42: LSAAK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
43: SAAKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
44: AKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
45: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
46: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
47: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
48: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
49: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
50: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
51: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK  
52: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK

(K){20}  
53: KKKKK KKKKKKKKKKKKKKKKKKK KX  
(K){20}  
54: KKKKK KKKKKKKKKKKKKKKKKKK X

AAU18206 ck: 9217 len: 68 1 Aau18206 Novel human DNA-binding protein #5

(R,K){20}

38: FLEPE KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

39: LPPEK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

40: FPEKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

41: PEKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

42: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

43: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

45: KKKKK KKKKKKKKKKKKKKKKKKK GXXF

AAU18208 ck: 8152 len: 79 1 Aau18208 Novel human DNA-binding protein #5

(R,K){20}

41: VRPEV RKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

42: RPRVR KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

43: PRVRK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

44: RVRRK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

45: VRKKK KKKKKKKKKKKKKKKKKKK GGRFR

AAU18237 ck: 285 len: 118 1 Aau18237 Novel human DNA-binding protein #8

(R,K){20}

98: EKHQK KKKKKKKKKKKKKKKKKKKR G

(R,K){20}

AAU18238 ck: 5509 len: 58 1 Aau18238 Novel human DNA-binding protein #8

(R,K){20}

36: FYEVC KKKKKKKKKKKKKKKKKKK KKK

(K){20}

37: YFVCK KKKKKKKKKKKKKKKKKKK KKK

(K){20}

38: FVCKK KKKKKKKKKKKKKKKKKKK K

1

(K){20}  
39: VCKKK KKKKKKKKKKKKKKKKKKK

AAU18239 ck: 9074 len: 66 1 Aau18239 Novel human DNA-binding protein

(R,K){20}

40: LVQCE KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

41: VQCEK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

42: QCEKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

43: CEKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

44: EKKKK KKKKKKKKKKKKKKKKKKK KKK

(K){20}

45: KKKKK KKKKKKKKKKKKKKKKKKK KKK

(K){20}

46: KKKKK KKKKKKKKKKKKKKKKKKK K

(K){20}

47: KKKKK KKKKKKKKKKKKKKKKKKK

AAU18240 ck: 8528 len: 150 1 Aau18240 Novel human DNA-binding protein

(R,K){20}

113: SRNTV KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

114: RNTVK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

115: NTVKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

116: TVKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

117: VKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

118: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

119: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

120: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

121: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

122: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

123: KKKKK KKKKKKKKKKKKKKKKKKK XXAVL

AAU18241 ck: 7676 len: 156 1 Aau18241 Novel human DNA-binding protein

(R,K){20}

(K){20}

1

108: KTWI KKKKKKKKKKKKKKKKKKK  
(K){20}  
109: TWIK KKKKKKKKKKKKKKKKKKK  
(K){20}  
110: TWIK KKKKKKKKKKKKKKKKKKK  
(K){20}  
111: WIKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
112: IKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
113: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
114: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
115: KKKK KKKKKKKKKKKKKKKKKKK  
(R,K){20}

AAU18242 ck: 1736 len: 40 1 Aau18242 Novel human DNA-binding protein #8

(R,K){20}  
(K){20}  
18: LPSL KKKKKKKKKKKKKKKKKKK  
(K){20}  
19: PSLK KKKKKKKKKKKKKKKKKKK  
(K){20}  
20: GSKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
21: SLKK KKKKKKKKKKKKKKKKKKK  
(K){20}

AAU18244 ck: 1109 len: 98 1 Aau18244 Novel human DNA-binding protein #9

(R,K){20}  
(K){20}  
53: QTKN KKKKKKKKKKKKKKKKKKK  
(K){20}  
54: TKNT KKKKKKKKKKKKKKKKKKK  
(K){20}  
55: KNTK KKKKKKKKKKKKKKKKKKK  
(K){20}  
56: NTKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
57: TKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
58: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
59: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
60: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
61: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
62: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

63: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
64: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
65: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
66: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
67: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
68: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

AAU18246 ck: 8102 len: 111 1 Aau18246 Novel human DNA-binding protei

(R,K){20}

78: EFHL KKKKKKKKKKKKKKKKKKK  
(K){20}

79: FHIL KKKKKKKKKKKKKKKKKKK  
(K){20}

80: HILK KKKKKKKKKKKKKKKKKKK  
(K){20}

81: ILKK KKKKKKKKKKKKKKKKKKK  
(K){20}

82: LKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

83: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

84: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

85: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

86: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

87: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

88: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

89: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

90: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

91: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

92: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

AAU18247 ck: 8102 len: 111 1 Aau18247 Novel human DNA-binding protei

(R,K){20}

78: EFHL KKKKKKKKKKKKKKKKKKK  
(K){20}

```

      (K){20}
79: FHILK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
80: HILKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
81: ILKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
82: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
83: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
84: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
85: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
86: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
87: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
88: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
89: KKKKK KKKKKKKKKKKKKKKKKKK KKK
      (K){20}
90: KKKKK KKKKKKKKKKKKKKKKKKK KK
      (K){20}
91: KKKKK KKKKKKKKKKKKKKKKKKK K
      (K){20}
92: KKKKK KKKKKKKKKKKKKKKKKKK

AAU18248  ck: 8319  len: 53  ! Aau18248 Novel human DNA-binding protein #9
      (R,K){20}
      (K){20}
13: RYFPP KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
14: YFPPK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
15: FPPKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
16: KPPKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
17: PPKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

```

```

      (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

AAU18250  ck: 7918  len: 80  ! Aau18250 Novel human DNA-binding protein
      (R,K){20}
      (K){20}
50: NVLTV KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
51: VLTVK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
52: LTVKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
53: TVKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
54: VKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

AAU18252  ck: 4882  len: 41  ! Aau18252 Novel human DNA-binding protein
      (R,K){20}
      (K){20}
8: FYCFP KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
9: YCFPP KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
10: CFPKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
11: FFKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
12: FKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}

```

13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
20: KKKKK KKKKKKKKKKKKKKKKKKK GX  
(K){20}

AAU18253 ck: 5469 len: 63 i Aau18253 Novel human DNA-binding protein #1

1  
30: ITCLL KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
(K){20}  
31: TCLKL KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
32: CLCLK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
33: LKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
34: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
35: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
36: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
37: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
38: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
39: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
40: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
41: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
42: KKKKK KKKKKKKKKKKKKKKKKKK XG  
(K){20}

AAU18254 ck: 5075 len: 52 i Aau18254 Novel human DNA-binding protein #1

1  
30: FIVVX KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
(K){20}

31: IVVKK KKKKKKKKKKKKKKKKKKK KK  
(K){20}  
32: VVKKK KKKKKKKKKKKKKKKKKKK K  
(K){20}  
33: VKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

AAU18255 ck: 5741 len: 47 i Aau18255 Novel human DNA-binding prote

1  
20: ILTTF KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
(K){20}  
21: LTTFK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
22: TTFKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
23: TFKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
24: FKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
27: KKKKK KKKKKKKKKKKKKKKKKKK X  
(K){20}

AAU18256 ck: 2868 len: 84 i Aau18256 Novel human DNA-binding protei

1  
53: KCTYE KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
(K){20}  
54: CTYEK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
55: TYEKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
56: YEKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
57: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
58: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
59: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
60: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
61: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
62: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

AAU18257 ck: 4686 len: 73 i Aau18257 Novel human DNA-binding protei

1

(R,K){20}

(K){20}

41: YLKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

42: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

43: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

45: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

50: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

51: KKKKK KKKKKKKKKKKKKKKKKKK KKK

(K){20}

52: KKKKK KKKKKKKKKKKKKKKKKKK KK

(K){20}

53: KKKKK KKKKKKKKKKKKKKKKKKK K

(K){20}

54: KKKKK KKKKKKKKKKKKKKKKKKK

AAU18258 ck: 6676 len: 74 1 Aau18258 Novel human DNA-binding protein #1

(R,K){20}

(K){20}

47: LPTPK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

48: RPTPK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

49: TPTPK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

50: PPKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

51: OKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

52: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

AAU18259 ck: 2283 len: 54 1 Aau18259 Novel human DNA-binding protein #1

(R,K){20}

(K){20}

32: TVPKK KKKKKKKKKKKKKKKKKKK KKK

1

(R,K){20}

(K){20}

45: SHLTD KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

46: HLTDK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

47: LTDDK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

48: TDDKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

49: DKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

50: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

51: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

52: KKKKK KKKKKKKKKKKKKKKKKKK KKK

(K){20}

53: KKKKK KKKKKKKKKKKKKKKKKKK KK

(K){20}

54: KKKKK KKKKKKKKKKKKKKKKKKK K

(K){20}

55: KKKKK KKKKKKKKKKKKKKKKKKK

AAU18262 ck: 5199 len: 84 1 Aau18262 Novel human DNA-binding protein

(R,K){20}

(K){20}

63: AKNAS KKKKKKKKKKKKKKKKKKK XG

AAU18263 ck: 7578 len: 31 1 Aau18263 Novel human DNA-binding protein

(R,K){20}

(K){20}

6: LTELK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

7: TELKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

8: ELEKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

9: LEKKK KKKKKKKKKKKKKKKKKKK KKK

(K){20}

10: EKKKK KKKKKKKKKKKKKKKKKKK KKK

(K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKK X



```
1
AAU18264 ck: 3915 len: 57 ! Aau18264 Novel human DNA-binding protein #1
(R,K){20}
(K){20}
31: KQLLL KKKKKKKKKKKKKKKKKKK KKKXG
(K){20}
32: QLLK KKKKKKKKKKKKKKKKKKK KXXG
(K){20}
33: LLLK KKKKKKKKKKKKKKKKKKK KXGGE
(K){20}
34: LLKK KKKKKKKKKKKKKKKKKKK XGGP

AAU18265 ck: 3679 len: 37 ! Aau18265 Novel human DNA-binding protein #1
(R,K){20}
(K){20}
15: ISPL KKKKKKKKKKKKKKKKKKK KXX
(K){20}
16: SPLK KKKKKKKKKKKKKKKKKKK KX
(K){20}
17: PLKK KKKKKKKKKKKKKKKKKKK X

AAU18266 ck: 657 len: 196 ! Aau18266 Novel human DNA-binding protein #1
(R,K){20}
(K){20}
169: FVFE KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
170: VVFE KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
171: XFEE KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
172: FEKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
173: EKKK KKKKKKKKKKKKKKKKKKK KKKX
(K){20}
174: KKKK KKKKKKKKKKKKKKKKKKK KXX
(K){20}
175: KKKK KKKKKKKKKKKKKKKKKKK KX
(K){20}
176: KKKK KKKKKKKKKKKKKKKKKKK X

AAU18267 ck: 4672 len: 57 ! Aau18267 Novel human DNA-binding protein #1
(R,K){20}
(K){20}
28: DTFE KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
29: KTFE KKKKKKKKKKKKKKKKKKK KKKXP
(K){20}
30: TFHK KKKKKKKKKKKKKKKKKKK KXXPG
(K){20}
31: FHKK KKKKKKKKKKKKKKKKKKK KXPG
(K){20}

1
32: HKKK KKKKKKKKKKKKKKKKKKK XPGG

AAU18268 ck: 9656 len: 66 ! Aau18268 Novel human DNA-binding protein #1
(R,K){20}
(K){20}
38: WVSV KKKKKKKKKKKKKKKKKKK KKKR
(K){20}
39: VISV KKKKKKKKKKKKKKKKKKK KKKRE
(K){20}
40: ISVK KKKKKKKKKKKKKKKKKKK KKKRE
(K){20}
41: SVKK KKKKKKKKKKKKKKKKKKK KKKEX
(K){20}
42: VKKK KKKKKKKKKKKKKKKKKKK KKKX
(K){20}
43: KKKK KKKKKKKKKKKKKKKKKKK EKK

AAU18270 ck: 4665 len: 34 ! Aau18270 Novel human DNA-binding protein #1
(R,K){20}
(K){20}
10: PELL KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
11: ELLK KKKKKKKKKKKKKKKKKKK KKK
(K){20}
12: LLK KKKKKKKKKKKKKKKKKKK KKK
(K){20}
13: LLK KKKKKKKKKKKKKKKKKKK KKK
(K){20}
14: LKK KKKKKKKKKKKKKKKKKKK K
(K){20}
15: KKKK KKKKKKKKKKKKKKKKKKK

AAU18271 ck: 7810 len: 64 ! Aau18271 Novel human DNA-binding protein #1
(R,K){20}
(K){20}
37: LKYE KKKKKKKKKKKKKKKKKKK KKKX
(K){20}
38: KYEW KKKKKKKKKKKKKKKKKKK KXXG
(K){20}
39: YEWK KKKKKKKKKKKKKKKKKKK KXXG
(K){20}
40: FWKK KKKKKKKKKKKKKKKKKKK KXXGP
(K){20}
41: WKKK KKKKKKKKKKKKKKKKKKK XGXP

AAU18272 ck: 269 len: 76 ! Aau18272 Novel human DNA-binding protein #1
(R,K){20}
(K){20}
37: TPRS KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
```

38: PSRAK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
39: SPRAK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
40: RAKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
41: AKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
42: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
43: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

AAU18273 ck: 8370 len: 45 1 Aau18273 Novel human DNA-binding protein #1  
(R,K){20}  
17: APRTQ KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
18: PRTOK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
19: KTOKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
20: TOKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
21: OKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
25: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

AAU18274 ck: 1663 len: 87 1 Aau18274 Novel human DNA-binding protein #1  
(R,K){20}  
36: KMSXX KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
37: WSSXX KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
38: SSXXX KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
39: SXKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
40: XKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
41: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

42: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
43: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
45: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
46: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
47: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
48: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
49: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
50: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
51: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
52: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
53: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
54: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
55: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
56: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
57: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
58: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
59: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
60: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
61: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
62: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
63: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

AAU18275 ck: 5607 len: 63 1 Aau18275 Novel human DNA-binding protein  
(R,K){20}  
26: MVELE KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}

```

27: VLEKK KKKKKKKKKKKKKKKKKKK KKKKK (K){20}
      (K){20}
28: ELEKK KKKKKKKKKKKKKKKKKKK KKKKK (K){20}
      (K){20}
29: LEKKK KKKKKKKKKKKKKKKKKKK KKKKK (K){20}
      (K){20}
30: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK (K){20}
      (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK (K){20}
      (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK (K){20}
      (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK (K){20}
      (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK (K){20}
      (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK (K){20}
      (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK (K){20}
      (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKK KKKGG (K){20}
      (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKK KKGGP (K){20}
      (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKK KGGPP (K){20}
      (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKK GGPP (K){20}
      (K){20}

AAU18276 ck: 5997 len: 58 i Aau18276 Novel human DNA-binding protein #1
      (R,K){20}
      (K){20}
28: RPTPK KKKKKKKKKKKKKKKKKKK KKKKK (K){20}
      (K){20}
29: PTPPK KKKKKKKKKKKKKKKKKKK XXXKK (K){20}
      (K){20}

AAU18277 ck: 5764 len: 74 i Aau18277 Novel human DNA-binding protein #1
      (R,K){20}
      (K){20}
40: EFLSA KKKKKKKKKKKKKKKKKKK KKKKK (K){20}
      (K){20}
41: FLSAK KKKKKKKKKKKKKKKKKKK KKKKK (K){20}
      (K){20}
42: LSAAK KKKKKKKKKKKKKKKKKKK KKKKK (K){20}
      (K){20}
43: SAKKK KKKKKKKKKKKKKKKKKKK KKKKK (K){20}
      (K){20}
44: AKKKK KKKKKKKKKKKKKKKKKKK KKKKK (K){20}
      (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK (K){20}
      (K){20}

```

```

46: KKKKK (K){20} KKKKK
47: KKKKK (K){20} KKKKK
48: KKKKK (K){20} KKKKK
49: KKKKK (K){20} KKKKK
50: KKKKK (K){20} KKKKK
51: KKKKK (K){20} KKKKK
52: KKKKK (K){20} KKKKK
53: KKKKK (K){20} KKKKK
54: KKKKK (K){20} KKKKK
55: KKKKK (K){20} KKKKK
56: KKKKK (K){20} KKKKK
57: KKKKK (K){20} KKKKK
58: KKKKK (K){20} KKKKK
59: KKKKK (K){20} KKKKK
60: KKKKK (K){20} KKKKK
61: KKKKK (K){20} KKKKK
62: KKKKK (K){20} KKKKK
63: KKKKK (K){20} KKKKK
64: KKKKK (K){20} KKKKK
65: KKKKK (K){20} KKKKK
66: KKKKK (K){20} KKKKK
67: KKKKK (K){20} KKKKK
68: KKKKK (K){20} KKKKK
69: KKKKK (K){20} KKKKK
70: KKKKK (K){20} KKKKK

```

71: KKKKK (K){20}  
 72: KKKKK (K){20}  
 73: KKKKK (K){20}  
 74: KKKKK (K){20}  
 75: KKKKK (K){20}

AAU18279 ck: 3164 len: 181 1 Aau18279 Novel human DNA-binding protein #1

154: TRKPE (R,K){20}  
 155: EKPEK (K){20}  
 156: KPEKK (K){20}  
 157: PEKKK (K){20}  
 158: EKKKK (K){20}  
 159: KKKKK (K){20}  
 160: KKKKK (K){20}  
 161: KKKKK (K){20}

AAU18280 ck: 7117 len: 55 1 Aau18280 Novel human DNA-binding protein #1

22: DDKKN (R,K){20}  
 23: DKKNK (K){20}  
 24: KKNKK (K){20}  
 25: KNNKK (K){20}  
 26: NKKKK (K){20}  
 27: KKKKK (K){20}

AAU18281 ck: 9316 len: 67 1 Aau18281 Novel human DNA-binding protein #1

36: FTLOT (R,K){20}

37: TLQTR (K){20}  
 38: LOTRK (K){20}  
 39: QTRKK (K){20}  
 40: TRKKK (K){20}  
 41: RKKKK (K){20}

AAM53862 ck: 5383 len: 86 1 Aam53862 Human brain expressed single ex

15: RRRRG (R,K){20}  
 16: RRGRG (R,K){20}  
 17: RRRGR (R,K){20}  
 18: RGRRR (R,K){20}  
 19: GRRRR (R,K){20}  
 20: RRRRR (R,K){20}  
 21: RRRRK (R,K){20}  
 22: RRRKK (R,K){20}  
 23: RRRKK (R,K){20}  
 24: RRRKK (R,K){20}  
 25: KKKKK (R,K){20}  
 26: KKKKK (R,K){20}  
 27: KKKKR (R,K){20}  
 28: KKKRK (R,K){20}  
 29: KKKKK (R,K){20}  
 30: KKKKK (R,K){20}  
 31: RKKKK (R,K){20}  
 32: KKKKK (R,K){20}



```
53: RRRR RRRRRRRRRRRRRRRR KEEE
    (R,K){20}
54: RRRR RRRRRRRRRRRRRRRR EEEE
    (K){20}
AAM5931 ck: 3607 len: 88 1 Aam5931 Human brain expressed single exon
    (R,K){20}
39: ERRE RRRRRRRRRRRRRRRR KKKK
    (K){20}
40: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
41: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
42: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
43: ERKK RRRRRRRRRRRRRRRR KKKK
    (K){20}
44: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
45: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
46: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
47: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
48: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
49: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
50: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
51: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
52: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
53: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
54: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
55: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
56: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
57: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
58: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
59: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
```

```
60: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
61: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
62: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
63: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
64: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
65: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
66: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
67: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
68: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
69: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
AAM56490 ck: 3937 len: 85 1 Aam56490 Human brain expressed single exon
    (R,K){20}
1: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
2: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
3: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
4: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
5: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
6: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
7: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
8: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
9: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
10: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
11: RRRR RRRRRRRRRRRRRRRR KKKK
    (K){20}
AAM58224 ck: 1560 len: 88 1 Aam58224 Human brain expressed single exon
    (R,K){20}
43: RRRR RRRRRRRRRRRRRRRR RRRR
    (K){20}
```

[illegible]

1  
7: TEREK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
8: EREKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
9: REKKK KKKKKKKKKKKKKKKKKKK NKKKK  
(R,K){20}  
AAM63533 ck: 2394 len: 57 ! Aam63533 Human brain expressed single e  
(R,K){20}  
20: EEEEG RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
21: EEEGR RRRRRRRRRRRRRRRRRRR RRRRG  
(R){20}  
22: EEEGR RRRRRRRRRRRRRRRRRRR RRRGR  
(R){20}  
23: EGRRR RRRRRRRRRRRRRRRRRRR RRRGR  
(R){20}  
24: GRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
25: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
AAM64090 ck: 4228 len: 24 ! Aam64090 Human brain expressed single e  
(R,K){20}  
1: RRRRRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
2: R RRRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
3: RR RRRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
4: RRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
AAM64863 ck: 3301 len: 52 ! Aam64863 Human brain expressed single e  
(R,K){20}  
12: KKKKN KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
16: NKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}

```

19: KKKKK (R,K){20}
20: KKKKK (R,K){20}
21: KKKKK (R,K){20}
22: KKKKK (R,K){20}
23: KKKKK (R,K){20}
24: KKKKK (R,K){20}
25: KKKKK (R,K){20}
26: KKKKK (R,K){20}
27: KKKKK (R,K){20}
28: KKKKK (R,K){20}
29: KKKKK (R,K){20}
30: KKKKK (R,K){20}
31: KKKKK (R,K){20}

AAM65355 ck: 4895 len: 51 1 Aam65355 Human brain expressed single exon
18: LKKKK (R,K){20}
19: KKKKK (R,K){20}
20: KKKKK (R,K){20}

AAM66249 ck: 5383 len: 86 1 Aam66249 Human bone marrow expressed probe
15: KKKKK (R,K){20}
16: KKKKK (R,K){20}
17: KKKKK (R,K){20}
18: KKKKK (R,K){20}
19: KKKKK (R,K){20}
20: KKKKK (R,K){20}

```

```

21: KKKKK (R,K){20}
22: KKKKK (R,K){20}
23: KKKKK (R,K){20}
24: KKKKK (R,K){20}
25: KKKKK (R,K){20}
26: KKKKK (R,K){20}
27: KKKKK (R,K){20}
28: KKKKK (R,K){20}
29: KKKKK (R,K){20}
30: KKKKK (R,K){20}
31: KKKKK (R,K){20}
32: KKKKK (R,K){20}
33: KKKKK (R,K){20}
34: KKKKK (R,K){20}
35: KKKKK (R,K){20}
36: KKKKK (R,K){20}
37: KKKKK (R,K){20}
38: KKKKK (R,K){20}
39: KKKKK (R,K){20}
40: KKKKK (R,K){20}
41: KKKKK (R,K){20}
42: KKKKK (R,K){20}
43: KKKKK (R,K){20}
44: KKKKK (R,K){20}

```



45: RRRR (R){20}  
RRRRRRRRRRRRRRRRRRR NKOTK

AAm67100 ck: 1334 len: 86 ! Aam67100 Human bone marrow expressed probe

57: EEEG (R,K){20}  
RRRRRRRRRRRRRRRRRRR KKKKK

58: EEEG (R,K){20}  
RRRRRRRRRRRRRRRRRRR KKKKK

59: EGGG (R,K){20}  
RRRRRRRRRRRRRRRRRRR KKKKK

60: EGGG (R,K){20}  
RRRRRRRRRRRRRRRRRRR KKKKK

61: GRRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR KKKKK

62: RRRK (R,K){20}  
RRRRRRRRRRRRRRRRRRR KKKKK

63: RRRK (R,K){20}  
RRRRRRRRRRRRRRRRRRR KKKK

64: RRRK (R,K){20}  
RRRRRRRRRRRRRRRRRRR KKK

65: KKKR (R,K){20}  
RRRRRRRRRRRRRRRRRRR KK

66: KKKR (R,K){20}  
RRRRRRRRRRRRRRRRRRR K

67: KKKR (K){20}  
RRRRRRRRRRRRRRRRRRR

AAm67180 ck: 9082 len: 167 ! Aam67180 Human bone marrow expressed probe

33: EGGG (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRGGG

34: EGGG (R,K){20}  
RRRRRRRRRRRRRRRRRRR RGGG

35: GGGG (R,K){20}  
RRRRRRRRRRRRRRRRRRR GGGG

AAm67717 ck: 2276 len: 89 ! Aam67717 Human bone marrow expressed probe

23: EEEE (R,K){20}  
RRRRRRRRRRRRRRRRRRR EEEK

46: KKEE (K){20}  
RRRRRRRRRRRRRRRRRRR KKEE

47: KKEE (K){20}  
RRRRRRRRRRRRRRRRRRR KKEE

48: EEEK (R,K){20}  
RRRRRRRRRRRRRRRRRRR EEEK

49: EEEK (R,K){20}  
RRRRRRRRRRRRRRRRRRR EEEE

1

AAm68007 ck: 1939 len: 130 ! Aam68007 Human bone marrow expressed pr

42: EGKE (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

43: GKKR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

44: RKEE (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

45: KERR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

46: ERRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

47: RRRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

48: RRRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

49: RRRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

50: RRRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

51: RRRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

52: RRRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

53: RRRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

54: RRRR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

AAm68298 ck: 3607 len: 88 ! Aam68298 Human bone marrow expressed pr

39: EKKR (R,K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

40: RKEE (K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

41: KKEE (K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

42: RKEE (K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

43: EKKK (K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

44: KKKK (K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

45: KKKK (K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

46: KKKK (K){20}  
RRRRRRRRRRRRRRRRRRR RRRR

47: KKKK (R){20}  
RRRRRRRRRRRRRRRRRRR RRRR

48: KKKKK (K){20} KKKKK  
 49: KKKKK (K){20} KKKKK  
 50: KKKKK (K){20} KKKKK  
 51: KKKKK (K){20} KKKKK  
 52: KKKKK (K){20} KKKKK  
 53: KKKKK (K){20} KKKKK  
 54: KKKKK (K){20} KKKKK  
 55: KKKKK (K){20} KKKKK  
 56: KKKKK (K){20} KKKKK  
 57: KKKKK (K){20} KKKKK  
 58: KKKKK (K){20} KKKKK  
 59: KKKKK (K){20} KKKKK  
 60: KKKKK (K){20} KKKKK  
 61: KKKKK (K){20} KKKKK  
 62: KKKKK (K){20} KKKKK  
 63: KKKKK (K){20} KKKKK  
 64: KKKKK (K){20} KKKKK  
 65: KKKKK (K){20} KKKKK  
 66: KKKKK (K){20} KKKKK  
 67: KKKKK (K){20} KKKKK  
 68: KKKKK (K){20} KKKKK  
 69: KKKKK (K){20} KKKKK  
 AAm68869 ck: 3937 len: 85 1 AAm68869 Human bone marrow expressed probe  
 1: KKKKK (K){20} KKKKK

2: KKKKK (K){20} KKKKK  
 3: KKKKK (K){20} KKKKK  
 4: KKKKK (K){20} KKKKK  
 5: KKKKK (K){20} KKKKK  
 6: KKKKK (K){20} KKKKK  
 7: KKKKK (K){20} KKKKK  
 8: KKKKK (K){20} KKKKK  
 9: KKKKK (K){20} KKKKK  
 10: KKKKK (K){20} KKKKK  
 11: KKKKK (K){20} KKKKK  
 AAm69569 ck: 2686 len: 71 1 AAm69569 Human bone marrow expressed probe  
 20: KKKKK (R,K){20} KKKKK  
 21: KKKKK (K){20} KKKKK  
 22: KKKKK (K){20} KKKKK  
 23: KKKKK (K){20} KKKKK  
 24: KKKKK (K){20} KKKKK  
 25: KKKKK (R,K){20} KKKKK  
 26: KKKKK (R,K){20} KKKKK  
 27: KKKKK (R,K){20} KKKKK  
 28: KKKKK (R,K){20} KKKKK  
 29: KKKKK (R,K){20} KKKKK  
 30: KKKKK (R,K){20} KKKKK  
 31: KKKKK (R,K){20} KKKKK  
 32: KKKKK (R,K){20} KKKKK

```
(R,K)[20]
33: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
34: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
35: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
36: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
37: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
38: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
39: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
40: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
41: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
42: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
43: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
AAM70678 ck: 1560 len: 88 1 Aam70678 Human bone marrow expressed probe
(R,K)[20]
43: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
AAM70678 ck: 1560 len: 88 1 Aam70678 Human bone marrow expressed probe
```

```
(R,K)[20]
49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
50: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
51: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
52: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
53: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
54: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
55: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
56: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
57: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
58: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
59: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
60: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
61: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
62: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
AAM70881 ck: 2324 len: 36 1 Aam70881 Human bone marrow expressed probe
(R,K)[20]
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
AAM73767 ck: 8343 len: 66 1 Aam73767 Human bone marrow expressed probe
(R,K)[20]
6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K)[20]
AAM76347 ck: 2394 len: 57 1 Aam76347 Human bone marrow expressed probe
(R,K)[20]
```

20: EEEEG RRRRRRRRRRRRRRRRRR RRRR  
(R,K){20}  
21: EEEGR RRRRRRRRRRRRRRRRRR RRRRG  
(R,K){20}  
22: EEEGR RRRRRRRRRRRRRRRRRR RRRGR  
(R,K){20}  
23: EGGRR RRRRRRRRRRRRRRRRRR RRRGR  
(R,K){20}  
24: GRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}  
25: RRRRR RRRRRRRRRRRRRRRRRR RRRRR  
(R,K){20}

AAM76911 ck: 4228 len: 24 ! Aam76911 Human bone marrow expressed probe

1: (R,K){20}  
(R,K){20}  
RRRRRRRRRRRRRRRRRR RRRRT  
(R,K){20}  
2: R RRRRRRRRRRRRRRRRRR RRT  
(R,K){20}  
3: RR RRRRRRRRRRRRRRRRRR RT  
(R,K){20}  
4: RRR RRRRRRRRRRRRRRRRRR T  
(R,K){20}

AAM78048 ck: 4895 len: 51 ! Aam78048 Human bone marrow expressed probe

1: (R,K){20}  
(R,K){20}  
18: LKPPM RRRRRRRRRRRRRRRR KLLTT  
(R,K){20}  
19: FKPMR KRRRRRRRRRRRRRRR KLLTT  
(R,K){20}  
20: KPMRK RRRRRRRRRRRRRRRR LTTT  
(R,K){20}

AAM82533 ck: 1736 len: 40 ! Aam82533 Human immune/haematopoietic antigen

1: (R,K){20}  
(K){20}  
18: LPSL KRRRRRRRRRRRRRRR KKK  
(K){20}  
19: PSLK KRRRRRRRRRRRRRRR KK  
(K){20}  
20: GSLK KRRRRRRRRRRRRRRR K  
(K){20}  
21: SLKK KRRRRRRRRRRRRRRR  
(K){20}

AAM85748 ck: 7503 len: 74 ! Aam85748 Human immune/haematopoietic antigen

1: (R,K){20}  
(K){20}  
45: SHLD KRRRRRRRRRRRRRRR KKKK  
(K){20}  
46: HLDK KRRRRRRRRRRRRRRR KKKK  
(K){20}

47: LTDK KRRRRRRRRRRRRRRR KKKK  
(K){20}

48: TDKK KRRRRRRRRRRRRRRR KKKK  
(K){20}

49: DKKK KRRRRRRRRRRRRRRR KKKK  
(K){20}

50: KKKK KRRRRRRRRRRRRRRR KKKK  
(K){20}

51: KKKK KRRRRRRRRRRRRRRR KKKK  
(K){20}

52: KKKK KRRRRRRRRRRRRRRR KKK  
(K){20}

53: KKKK KRRRRRRRRRRRRRRR KK  
(K){20}

54: KKKK KRRRRRRRRRRRRRRR K  
(K){20}

55: KKKK KRRRRRRRRRRRRRRR  
(K){20}

AAM90546 ck: 6676 len: 74 ! Aam90546 Human immune/haematopoietic antigen

1: (R,K){20}  
(K){20}

47: LRTQ KRRRRRRRRRRRRRRR KKKK  
(K){20}

48: RTQK KRRRRRRRRRRRRRRR KKKK  
(K){20}

49: TQKK KRRRRRRRRRRRRRRR KKKK  
(K){20}

50: FQKK KRRRRRRRRRRRRRRR KKKK  
(K){20}

51: QKKK KRRRRRRRRRRRRRRR KKK  
(K){20}

52: KKKK KRRRRRRRRRRRRRRR XKK  
(K){20}

AAM90618 ck: 5691 len: 108 ! Aam90618 Human immune/haematopoietic antigen

1: (R,K){20}  
(K){20}

78: VRPL KRRRRRRRRRRRRRRR KKKK  
(K){20}

79: RPCL KRRRRRRRRRRRRRRR KKKK  
(K){20}

80: PCLK KRRRRRRRRRRRRRRR KKKK  
(K){20}

81: CLKK KRRRRRRRRRRRRRRR KKKK  
(K){20}

82: LKKK KRRRRRRRRRRRRRRR KKKK  
(K){20}

83: KKKK KRRRRRRRRRRRRRRR KKKK  
(K){20}

84: KKKK KRRRRRRRRRRRRRRR KKKK  
(K){20}



87: KKKK (K){20} KKKK  
88: KKKK (K){20} KKKK  
89: KKKK (K){20} KKKK  
90: KKKK (K){20} KKKK  
91: KKKK (K){20} K  
92: KKKK (K){20} KKKK  
AAM9193 ck: 8285 len: 55 ! Aam9193 Human digestive system antigen SEQ  
(R,K){20}  
33: PTPRP KKKK (K){20} KKKK  
34: PTPRP KKKK (K){20} GK  
AAM92433 ck: 1663 len: 87 ! Aam92433 Human digestive system antigen SEQ  
(R,K){20}  
36: KWSX KKKK (K){20} KKKK  
37: WSSX KKKK (K){20} KKKK  
38: SSXK KKKK (K){20} KKKK  
39: SXKX KKKK (K){20} KKKK  
40: XKKK KKKK (R){20} KKKK  
41: KKKK KKKK (K){20} KKKK  
42: KKKK KKKK (R){20} KKKK  
43: KKKK KKKK (K){20} KKKK  
44: KKKK KKKK (K){20} KKKK  
45: KKKK KKKK (K){20} KKKK  
46: KKKK KKKK (K){20} KKKK  
47: KKKK KKKK (K){20} KKKK  
48: KKKK KKKK (K){20} KKKK  
(K){20}

49: KKKK KKKK KKKK KKKK KKKK  
50: KKKK (K){20} KKKK  
51: KKKK (K){20} KKKK  
52: KKKK (K){20} KKKK  
53: KKKK (K){20} KKKK  
54: KKKK (K){20} KKKK  
55: KKKK (K){20} KKKK  
56: KKKK (K){20} KKKK  
57: KKKK (K){20} KKKK  
58: KKKK (K){20} KKKK  
59: KKKK (K){20} KKKK  
60: KKKK (K){20} KKKK  
61: KKKK (K){20} KKKK  
62: KKKK (K){20} KKKK  
63: KKKK (R,K){20} KKKK  
AaO0092 ck: 9065 len: 113 ! AaO0092 Human polypeptide SEQ ID NO 139  
(R,K){20}  
91: CLGCL KKKK (K){20} KKF  
92: LGCLK KKKK (K){20} FF  
AaO00222 ck: 2916 len: 132 ! AaO00222 Human polypeptide SEQ ID NO 141  
(R,K){20}  
28: XPLPP KKKK (K){20} GPPK  
AaO00232 ck: 1000 len: 102 ! AaO00232 Human polypeptide SEQ ID NO 141  
(R,K){20}  
29: RDCEP KKKK (K){20} KKKK  
30: DCEPK KKKK (K){20} KKKK  
31: CPEPK KKKK (K){20} KKKK

32: FFKKK (R){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 33: FKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 34: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 35: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 36: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 37: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 38: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 39: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 40: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 41: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 42: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 43: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 44: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 45: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 46: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 47: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 48: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 49: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 50: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 51: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 52: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 53: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 54: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 55: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

AA000291 ck: 8100 len: 124 1 Aa000291 Human polypeptide SEQ ID NO 141  
 (R,K){20}  
 22: QHFCM (K){20} KKKKKKKKKKKKKKKKKKK FFKKK  
 AA000439 ck: 6396 len: 122 1 Aa000439 Human polypeptide SEQ ID NO 141  
 (R,K){20}  
 23: CLWLK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 24: LMLVK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 25: WLVRK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 26: LVKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 27: VKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 28: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 29: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 30: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 31: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 32: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 33: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 34: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 35: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 36: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 37: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 38: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 39: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 40: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 41: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 42: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 43: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK





```

51: KKKK (K)[20]
52: KKKK (K)[20]
53: KKKK (K)[20]
54: KKKK (K)[20]
AA002186 ck: 4844 len: 57 1 Aa002186 Human polypeptide SEQ ID NO 16078.
(R,K)[20]
11: HCCLL KKKKKKKKKKKKKKKKKKK KKKK
(K)[20]
12: CCLLK KKKKKKKKKKKKKKKKKKK KKKK
(K)[20]
AA002310 ck: 704 len: 137 1 Aa002310 Human polypeptide SEQ ID NO 16202.
(R,K)[20]
22: HSLNL KKKKKKKKKKKKKKKKKKK GCGVK
(R,K)[20]
1 AA002477 ck: 2018 len: 112 1 Aa002477 Human polypeptide SEQ ID NO 16369.
(R,K)[20]
83: LASAV KKKKKKKKKKKKKKKKKKK KKKK
(K)[20]
84: ASAVK KKKKKKKKKKKKKKKKKKK KKKK
(K)[20]
85: SAVKK KKKKKKKKKKKKKKKKKKK KKKK
(K)[20]
86: AVKKK KKKKKKKKKKKKKKKKKKK KKKK
(K)[20]
87: VKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K)[20]
88: KKKK (K)[20]
89: KKKK (K)[20]
90: KKKK (K)[20]
91: KKKK (K)[20]
92: KKKK (K)[20]
93: KKKK (K)[20]
AA002733 ck: 2860 len: 132 1 Aa002733 Human polypeptide SEQ ID NO 16625.
(R,K)[20]
86: FFFSL KKKKKKKKKKKKKKKKKKK KKKK
(K)[20]
```

```

87: FFSLK KKKKKKKKKKKKKKKKKKK KKKK
(K)[20]
88: FSLKK KKKKKKKKKKKKKKKKKKK KKKK
(K)[20]
89: SLKKK KKKKKKKKKKKKKKKKKKK KKKK
(K)[20]
90: LKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K)[20]
91: KKKK (K)[20]
92: KKKK (K)[20]
93: KKKK (K)[20]
94: KKKK (K)[20]
95: KKKK (K)[20]
96: KKKK (K)[20]
97: KKKK (K)[20]
98: KKKK (K)[20]
99: KKKK (K)[20]
100: KKKK (K)[20]
101: KKKK (R,K)[20]
AA002946 ck: 3649 len: 126 1 Aa002946 Human polypeptide SEQ ID NO 16
(R,K)[20]
25: DEATS KKKKKKKKKKKKKKKKKKK KKKK
(K)[20]
26: EATSK KKKKKKKKKKKKKKKKKKK KKKK
(K)[20]
27: ATSKK KKKKKKKKKKKKKKKKKKK KKKK
(K)[20]
28: TSKKK KKKKKKKKKKKKKKKKKKK KKKK
(K)[20]
29: SKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K)[20]
30: KKKK (K)[20]
31: KKKK (K)[20]
32: KKKK (K)[20]
```

33: KKKKK (K){20}  
34: KKKKK (K){20}  
35: KKKKK (K){20}  
36: KKKKK (K){20}  
37: KKKKK (K){20}  
38: KKKKK (K){20}

AAO02961 ck: 4320 len: 83 ! Aao02961 Human polypeptide SEQ ID NO 16853.

22: RWFSS (R,K){20}  
23: MFSKK (K){20}

AAO03006 ck: 5212 len: 102 ! Aao03006 Human polypeptide SEQ ID NO 16898.

59: HDFFP (R,K){20}  
60: DFFPK (K){20}  
61: FFFPK (K){20}  
62: FPKKK (K){20}  
63: PFKKK (K){20}  
64: KFKKK (K){20}  
65: KKKKK (K){20}  
66: KKKKK (K){20}  
67: KKKKK (K){20}

AAO03024 ck: 6865 len: 64 ! Aao03024 Human polypeptide SEQ ID NO 16916.

15: SPAKA (R,K){20}  
16: PAKAR (K){20}  
17: AKARK (R,K){20}

AAO03113 ck: 5587 len: 60 ! Aao03113 Human polypeptide SEQ ID NO 170

33: EINML (R,K){20}  
34: INMLK (K){20}  
35: NMLKK (K){20}  
36: MLKKK (K){20}  
37: LKKKK (R,K){20}

AAO03123 ck: 6627 len: 108 ! Aao03123 Human polypeptide SEQ ID NO 170

33: RTWRX (R,K){20}  
34: TWRXX (K){20}  
35: WRXXK (K){20}  
36: RXKKK (K){20}  
37: XXKKK (K){20}  
38: KKKKK (K){20}  
39: KKKKK (K){20}  
40: KKKKK (K){20}  
41: KKKKK (K){20}  
42: KKKKK (K){20}  
43: KKKKK (R,K){20}

AAO03132 ck: 3903 len: 116 ! Aao03132 Human polypeptide SEQ ID NO 170

1: KKKKK (R,K){20}  
2: KKKKK (K){20}  
3: KKKKK (K){20}  
4: KKKKK (K){20}

5: KKKK KKKKKKKKKKKKKKKKKK RGGPP  
(R,K){20}  
6: KKKKK KKKKKKKKKKKKKKKKKR GGPPK  
AAO03152 ck: 7891 len: 35 1 Aao03152 Human polypeptide SEQ ID NO 17044.  
(R,K){20}  
13: LIXYT RKKKKKKKKKKKKKKKKK RGG  
(R,K){20}  
14: IXYTR KKKKKKKKKKKKKKKKKR GG  
AAO03168 ck: 7233 len: 67 1 Aao03168 Human polypeptide SEQ ID NO 17060.  
(R,K){20}  
17: ELIMP KKKKKKKKKKKKKKKKKK KRGA  
(K){20}  
18: LMPK KKKKKKKKKKKKKKKKKK RGGAP  
(K){20}  
19: LMPK KKKKKKKKKKKKKKKKKR GGAP  
AAO03243 ck: 411 len: 95 1 Aao03243 Human polypeptide SEQ ID NO 17135.  
(R,K){20}  
51: LHAIV KKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
52: LHAIV KKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
53: HAVK KKKKKKKKKKKKKKKKKK EKKK  
AAO03273 ck: 2981 len: 88 1 Aao03273 Human polypeptide SEQ ID NO 17165.  
(R,K){20}  
43: IXYL KKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
44: XYLL KKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
45: YLKK KKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
46: LKKK KKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
47: LKKK KKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
48: KKKK KKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
49: KKKK KKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
50: KKKK KKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
51: KKKK KKKKKKKKKKKKKKKKKK KKKK  
(K){20}

52: KKKK KKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
53: KKKK KKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
54: KKKK KKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
55: KKKK KKKKKKKKKKKKKKKKKK KKP  
(K){20}  
56: KKKK KKKKKKKKKKKKKKKKK KPPK  
(K){20}  
57: KKKK KKKKKKKKKKKKKKKKK PTPK  
AAO03277 ck: 2981 len: 89 1 Aao03277 Human polypeptide SEQ ID NO 17.  
(R,K){20}  
1: KKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
2: K KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
3: KK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
4: KK KKKKKKKKKKKKKKKKK KKKK  
(R,K){20}  
5: KKK KKKKKKKKKKKKKKKK KKKK  
AAO03278 ck: 6777 len: 74 1 Aao03278 Human polypeptide SEQ ID NO 17.  
(R,K){20}  
47: RAVQ KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
48: AVAQ KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
49: VAQK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
50: AQKK KKKKKKKKKKKKKKKKK KKKK  
AAO03284 ck: 9650 len: 115 1 Aao03284 Human polypeptide SEQ ID NO 17.  
(R,K){20}  
22: KALV KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
23: ALVK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
24: LVSK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
25: VSKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
26: SKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
27: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

```
(R,K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKK RGAL
(R,K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKR GGALK

AA0328 ck: 1109 len: 117 1 Aa0328 Human polypeptide SEQ ID NO 17180.

(R,K){20}
31: SCSYP KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
32: CSYPR KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
33: SYPRK KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
34: YPRKK KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
35: PRKKK KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKR KKKK
(R,K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKR KKKK
(R,K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKR KKKK
(R,K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKR KKKK
(R,K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKR KKKK
(R,K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKR KKKK
(R,K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKR KKKK
(R,K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKR KKKK
```

```
(R,K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKR KTXGG
(R,K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKR TXGGG

AA03375 ck: 693 len: 54 1 Aa03375 Human polypeptide SEQ ID NO 172

(R,K){20}
25: TXLEB KKKKKKKKKKKKKKKKKKKR KKKAP
(K){20}
26: XLEBK KKKKKKKKKKKKKKKKKKKR KRAPG
(K){20}
27: LEEKK KKKKKKKKKKKKKKKKKKKR KAPGG
(K){20}
28: EEEKK KKKKKKKKKKKKKKKKKKKR APGGG

AA03459 ck: 5217 len: 58 1 Aa03459 Human polypeptide SEQ ID NO 173

(R,K){20}
28: XMCVF KKKKKKKKKKKKKKKKKKKR KDPPK
(K){20}
29: MCVFR KKKKKKKKKKKKKKKKKKKR DPEKK
(K){20}

AA03475 ck: 315 len: 138 1 Aa03475 Human polypeptide SEQ ID NO 173

(R,K){20}
44: LLXLT KKKKKKKKKKKKKKKKKKKR KKKKK
(K){20}
45: LXLTK KKKKKKKKKKKKKKKKKKKR KKKKG
(K){20}
46: XLTKK KKKKKKKKKKKKKKKKKKKR KKKGG
(K){20}
47: LTKKK KKKKKKKKKKKKKKKKKKKR KKGPP
(K){20}
48: TKKKK KKKKKKKKKKKKKKKKKKKR KGGPR
(K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKR GGPRG
(K){20}

AA03615 ck: 7613 len: 128 1 Aa03615 Human polypeptide SEQ ID NO 175

(R,K){20}
28: PLPRX KKKKKKKKKKKKKKKKKKKR KKTAA
(K){20}
29: LPXXK KKKKKKKKKKKKKKKKKKKR KTAAG
(K){20}
30: PRXXK KKKKKKKKKKKKKKKKKKKR TAAAG
(K){20}

AA03623 ck: 3884 len: 116 1 Aa03623 Human polypeptide SEQ ID NO 175

(R,K){20}
(K){20}
```

```

77: TPSPA KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
78: PSRAK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
79: SRAKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
80: RAKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
81: AKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
82: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
83: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
84: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
85: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
86: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
87: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
88: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}

AA003700 ck: 1279 len: 43 ! Aa03700 Human polypeptide SEQ ID NO.17592
      (R,K){20}
      (K){20}
19: YSQRK KKKKKKKKKKKKKKKKKKKKK GGGPF

AA003703 ck: 7016 len: 113 ! Aa03703 Human polypeptide SEQ ID NO.17595
      (R,K){20}
      (K){20}
58: IWDAA KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
59: WDAEK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
60: DAERK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
61: AEKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
62: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKKK SPGGA
      (K){20}

AA003766 ck: 8808 len: 81 ! Aa03766 Human polypeptide SEQ ID NO.17658

```

1 (R,K){20}  
(K){20}  
49: TTAC KKKKKKKKKKKKKKKKKKK KKKK  
50: TTAC KKKKKKKKKKKKKKKKKKK KKKK (K){20}  
51: TACK KKKKKKKKKKKKKKKKKKK KKKK (K){20}  
52: ACKK KKKKKKKKKKKKKKKKKKK KKKK (K){20}  
53: CKKK KKKKKKKKKKKKKKKKKKK KKKK (K){20}  
54: KKKK KKKKKKKKKKKKKKKKKKK KKKG (K){20}  
55: KKKK KKKKKKKKKKKKKKKKKKK KKGK (K){20}  
56: KKKK KKKKKKKKKKKKKKKKKKK KGGG (K){20}  
57: KKKK KKKKKKKKKKKKKKKKKKK GGGA (K){20}

AA03841 ck: 8734 len: 100 ! Aa03841 Human polypeptide SEQ ID NO 17.  
(R,K){20}  
36: KQHP KKKKKKKKKKKKKKKKKKK KKKK (K){20}  
37: QHPK KKKKKKKKKKKKKKKKKKK KKKK (K){20}  
38: HYPK KKKKKKKKKKKKKKKKKKK KKKK (K){20}  
39: YPKK KKKKKKKKKKKKKKKKKKK KKKK (K){20}  
40: PKKK KKKKKKKKKKKKKKKKKKK KKKK (K){20}  
41: KKKK KKKKKKKKKKKKKKKKKKK KKRT (K){20}  
42: KKKK KKKKKKKKKKKKKKKKKKK KKRT (K){20}  
43: KKKK KKKKKKKKKKKKKKKKKKK KRTQ (K){20}  
44: KKKK KKKKKKKKKKKKKKKKKKK RTQK (K){20}  
45: KKKK KKKKKKKKKKKKKKKKKKK TQKK (R,K){20}

AA03906 ck: 4312 len: 100 ! Aa03906 Human polypeptide SEQ ID NO 17.  
(R,K){20}  
34: NKON KKKKKKKKKKKKKKKKKKK KKKK (K){20}  
35: KONK KKKKKKKKKKKKKKKKKKK KKKK (K){20}

```

36: ONQKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
37: NQKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
38: OKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}

AAO03967 ck: 7917 len: 53 ! Aao03967 Human polypeptide SEQ ID NO 17859
    (R,K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
17: LKSKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
18: KSKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
19: KSKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
20: SKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}

AAO04512 ck: 4277 len: 62 ! Aao04512 Human polypeptide SEQ ID NO 18404
    (R,K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}

```

```

1
AAO04619 ck: 8654 len: 39 ! Aao04619 Human polypeptide SEQ ID NO 185
    (R,K){20}
13: LIIFL KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
14: ITFLK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
15: TFLKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
16: FLKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
17: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}

AAO04644 ck: 2038 len: 70 ! Aao04644 Human polypeptide SEQ ID NO 185
    (R,K){20}
23: AKPPT KKKKKKKKKKKKKKKKKKK KKKKK
    (R,K){20}

AAO04645 ck: 7825 len: 118 ! Aao04645 Human polypeptide SEQ ID NO 185
    (R,K){20}
46: IKSPK KKKKKKKKKKKKKKKKKKK KKKKK
    (R,K){20}
47: KSPFK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
48: SFLKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
49: FLKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
50: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}

```

58: KKKKK KKKKKKKKKKKKKKKKK KKKKP  
(K){20}  
59: KKKKK KKKKKKKKKKKKKKKKK KKKPR  
(K){20}  
60: KKKKK KKKKKKKKKKKKKKKKK KKKPRG  
(K){20}  
61: KKKKK KKKKKKKKKKKKKKKKK KPRGG  
(K){20}  
62: KKKKK KKKKKKKKKKKKKKKKK PRGGC  
(K){20}

AA004647 ck: 4805 len: 58 1 Aa004647 Human polypeptide SEQ ID NO 18539.  
(R,K){20}  
(K){20}

28: KPTRP KKKKKKKKKKKKKKKKK KKKKK  
(K){20}

29: PTRPK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}

30: TRPK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}

31: RPKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}

32: PKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}

33: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}

34: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}

35: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}

36: KKKKK KKKKKKKKKKKKKKKKK KKK  
(K){20}

37: KKKKK KKKKKKKKKKKKKKKKK KS  
(K){20}

38: KKKKK KKKKKKKKKKKKKKKKK S  
(K){20}

AA004674 ck: 2036 len: 33 1 Aa004674 Human polypeptide SEQ ID NO 18566.  
(R,K){20}  
(K){20}

9: WCYIT KKKKKKKKKKKKKKKKK KKAAG  
(K){20}

10: CYITK KKKKKKKKKKKKKKKKK KAAG  
(K){20}

11: YITPK KKKKKKKKKKKKKKKKK ARG  
(K){20}

AA004679 ck: 7190 len: 31 1 Aa004679 Human polypeptide SEQ ID NO 18571.  
(R,K){20}  
(K){20}

8: CWPTQ KKKKKKKKKKKKKKKKK KDGG  
(K){20}

9: WPTQK KKKKKKKKKKKKKKKKK DRG  
(K){20}

AA004682 ck: 5605 len: 60 1 Aa004682 Human polypeptide SEQ ID NO 185.  
(R,K){20}  
(K){20}

2: L KKKKKKKKKKKKKKKKK KKKKK  
(K){20}

3: LK KKKKKKKKKKKKKKKKK KKQKD  
(K){20}

4: LKK KKKKKKKKKKKKKKKKK KQKDL  
(K){20}

5: LKKK KKKKKKKKKKKKKKKKK QKDLK  
(K){20}

AA004690 ck: 7157 len: 81 1 Aa004690 Human polypeptide SEQ ID NO 186.  
(R,K){20}  
(K){20}

1: KKKKKKKKKKKKKKKKK KKKKK  
(K){20}

2: K KKKKKKKKKKKKKKKKK KKKKK  
(K){20}

3: KK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}

4: KKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}

5: KKKK KKKKKKKKKKKKKKKKK KKKPK  
(K){20}

6: KKKKK KKKKKKKKKKKKKKKKK KKKPK  
(K){20}

7: KKKKK KKKKKKKKKKKKKKKKK KPKKK  
(K){20}

8: KKKKK KKKKKKKKKKKKKKKKK PKKKK  
(K){20}

AA004715 ck: 6984 len: 35 1 Aa004715 Human polypeptide SEQ ID NO 187.  
(R,K){20}  
(K){20}

8: LGSKD KKKKKKKKKKKKKKKKK KKAKK  
(K){20}

9: GSKDK KKKKKKKKKKKKKKKKK KAKKK  
(K){20}

10: SKDKK KKKKKKKKKKKKKKKKK AKKKD  
(K){20}

AA004743 ck: 9412 len: 54 1 Aa004743 Human polypeptide SEQ ID NO 188.  
(R,K){20}  
(K){20}

21: KINKL KKKKKKKKKKKKKKKKK KRAAA  
(K){20}

22: INKLK KKKKKKKKKKKKKKKKK RAAAR  
(K){20}

23: NKLKK KKKKKKKKKKKKKKKKK AAARD  
(K){20}

AA004747 ck: 8399 len: 39 1 Aa004747 Human polypeptide SEQ ID NO 189.  
(K){20}

1

(R,K){20}  
13: RTGFV KKKKKKKKKKKKKKKKK KRGKG  
(K){20}  
14: TGFVK KKKKKKKKKKKKKKKKK KRGKG  
(K){20}  
15: GFVKK KKKKKKKKKKKKKKKKK RGGGF  
(R,K){20}  
16: FVKKK KKKKKKKKKKKKKKKKK GGGF

1

AA004752 ck: 1021 len: 71 1 Aa004752 Human polypeptide SEQ ID NO 18644.  
(R,K){20}  
19: QEGL KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
20: EGGLE KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
21: OGLKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
22: GLKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
23: LKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
24: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
25: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
26: KKKKK KKKKKKKKKKKKKKKKK KGGGL  
(K){20}  
27: KKKKK KKKKKKKKKKKKKKKKK GGGGL

1

AA004755 ck: 5521 len: 59 1 Aa004755 Human polypeptide SEQ ID NO 18647.  
(R,K){20}  
31: INSLE KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
32: NSLEK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
33: SLEKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
34: LEKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
35: EKRRK KKKKKKKKKKKKKKKKK KKKKT  
(K){20}  
36: KKKKK KKKKKKKKKKKKKKKKK KKKKT  
(K){20}  
37: KKKKK KKKKKKKKKKKKKKKKK KKT  
(K){20}  
38: KKKKK KKKKKKKKKKKKKKKKK KT

(K){20}  
39: KKKKK KKKKKKKKKKKKKKKKK T

1

AA004756 ck: 6952 len: 26 1 Aa004756 Human polypeptide SEQ ID NO 186  
(R,K){20}  
4: FFY KKKKKKKKKKKKKKKKK SSS  
(K){20}

1

AA004758 ck: 7836 len: 115 1 Aa004758 Human polypeptide SEQ ID NO 186  
(R,K){20}  
7: PFYOL KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
8: FYOLK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
9: YOLKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: OLKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: LKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
19: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
20: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
21: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
22: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
23: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
24: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
25: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
26: KKKKK KKKKKKKKKKKKKKKKK KKKKK



27: KKKK (K){20} KKKK  
28: KKKK (K){20} KKKK  
29: KKKK (K){20} KKKK  
30: KKKK (K){20} KKKK  
31: KKKK (K){20} KKKK  
32: KKKK (K){20} KKKK  
33: KKKK (K){20} KKKK  
34: KKKK (K){20} KKKK  
35: KKKK (K){20} KKKK  
36: KKKK (K){20} KKKK  
37: KKKK (K){20} KKKK  
38: KKKK (K){20} KKKK  
39: KKKK (K){20} KKKK  
40: KKKK (K){20} KKKK  
41: KKKK (K){20} KKKK  
42: KKKK (K){20} KKKK  
43: KKKK (K){20} KKKK  
44: KKKK (K){20} KKKK  
45: KKKK (K){20} KKKK  
46: KKKK (K){20} KKKK  
47: KKKK (K){20} KKKK  
48: KKKK (K){20} KKKK  
49: KKKK (K){20} KKKK  
50: KKKK (K){20} KKKK

1  
AA004764 ck: 2035 len: 54 ! Aa004764 Human polypeptide SEQ ID NO 186  
9: GDSSL (R,K){20} KKKK  
10: DSSLK (K){20} KKKK  
11: SSLK (K){20} KKKK  
12: SLKK (K){20} KKKK  
13: LKKK (K){20} KKKK  
14: KKKK (K){20} KKKK  
AA004802 ck: 1223 len: 105 ! Aa004802 Human polypeptide SEQ ID NO 186  
34: FFSRQ (R,K){20} KKKK  
AA004854 ck: 4121 len: 58 ! Aa004854 Human polypeptide SEQ ID NO 187  
13: TPFRK (R,K){20} KKKK  
14: PFRK (K){20} KKKK  
15: FRKK (K){20} KKKK  
16: FRKK (K){20} KKKK  
17: AKKK (K){20} KKKK  
18: KKKK (K){20} KKKK  
19: KKKK (K){20} KKKK  
20: KKKK (K){20} KKKK  
21: KKKK (K){20} KKKK  
22: KKKK (K){20} KKKK  
23: KKKK (K){20} KKKK  
24: KKKK (K){20} KKKK  
25: KKKK (K){20} KKKK

26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
27: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
28: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
29: KKKKK KKKKKKKKKKKKKKKKKKK KKRGG  
(K){20}  
30: KKKKK KKKKKKKKKKKKKKKKKKK KRGGN  
(K){20}  
31: KKKKK KKKKKKKKKKKKKKKKKKK RGGNF  
(K){20}  
32: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKR GGNFK

AA004856 ck: 4861 len: 135 1 Aa004856 Human polypeptide SEQ ID NO 18748.

41: LTLTT KKKKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
(K){20}  
42: TLTTK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
43: LTTKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
44: TTKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
45: TTKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
50: KKKKK KKKKKKKKKKKKKKKKKKK KKRGG  
(K){20}  
51: KKKKK KKKKKKKKKKKKKKKKKKK KKGCG  
(K){20}  
52: KKKKK KKKKKKKKKKKKKKKKKKK KGGGG  
(K){20}  
53: KKKKK KKKKKKKKKKKKKKKKKKK GGGGP  
(K){20}

AA004872 ck: 3224 len: 93 1 Aa004872 Human polypeptide SEQ ID NO 18764.

(R,K){20}  
(K){20}  
8: FLYKL KKKKKKKKKKKKKKKKKKK KKRGG  
(K){20}  
9: LYKLK KKKKKKKKKKKKKKKKKKK KKRGG  
(K){20}

(K){20}  
10: YKLKK KKKKKKKKKKKKKKKKKKK KRGGP  
(K){20}  
11: KLKKK KKKKKKKKKKKKKKKKKKK RGGPL  
(R,K){20}  
12: LKKKK KKKKKKKKKKKKKKKKKKK GGPLK  
(K){20}

AA004873 ck: 7719 len: 66 1 Aa004873 Human polypeptide SEQ ID NO 187

(R,K){20}  
(K){20}  
21: SFLLK KKKKKKKKKKKKKKKKKKK KGGGP  
(K){20}  
22: FLLEK KKKKKKKKKKKKKKKKKKK GGGPL  
(K){20}

AA004874 ck: 5753 len: 58 1 Aa004874 Human polypeptide SEQ ID NO 187

(R,K){20}  
(K){20}  
10: SEKLP KKKKKKKKKKKKKKKKKKK KKGKK  
(K){20}  
11: EKLPK KKKKKKKKKKKKKKKKKKK KGKKI  
(K){20}  
12: KLPPK KKKKKKKKKKKKKKKKKKK GKIKI  
(K){20}

AA004881 ck: 8841 len: 115 1 Aa004881 Human polypeptide SEQ ID NO 187

(R,K){20}  
(K){20}  
24: ITPHP KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
25: TPHPK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
26: PPHPK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
27: HPKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
28: PKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
29: KKKKK KKKKKKKKKKKKKKKKKKK KKGGA  
(K){20}  
30: KKKKK KKKKKKKKKKKKKKKKKKK KGAL  
(K){20}  
31: KKKKK KKKKKKKKKKKKKKKKKKK GGALK  
(K){20}

AA004917 ck: 3999 len: 103 1 Aa004917 Human polypeptide SEQ ID NO 188

(R,K){20}  
(K){20}  
1: KKKKKKKKKKKKKKKKKKKKK GGGPL  
(K){20}

AA004928 ck: 23 len: 43 1 Aa004928 Human polypeptide SEQ ID NO 188

(R,K){20}  
(K){20}  
19: FKKEK KKKKKKKKKKKKKKKKKKK GGGGF  
(K){20}

1  
AA004969 ck: 4345 len: 57 ! Aa004969 Human polypeptide SEQ ID NO 18861.  
(R,K){20}

26: LMGTS KKKKKKKKKKKKKKKKKKK  
(K){20}

27: MGTSK KKKKKKKKKKKKKKKKKKK  
(K){20}

28: GTSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

29: TSRSK KKKKKKKKKKKKKKKKKKK  
(K){20}

30: SKRSK KKKKKKKKKKKKKKKKKKK  
(K){20}

31: KRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

32: KRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

33: KRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

34: KRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

35: KRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

36: KRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

37: KRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

38: KRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

1  
AA005000 ck: 5847 len: 38 ! Aa005000 Human polypeptide SEQ ID NO 18892.  
(R,K){20}

5: SGRS KKKKKKKKKKKKKKKKKKK  
(K){20}

6: SGRS KKKKKKKKKKKKKKKKKKK  
(K){20}

7: GRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

8: RSRSK KKKKKKKKKKKKKKKKKKK  
(K){20}

9: SKRSK KKKKKKKKKKKKKKKKKKK  
(K){20}

10: KRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

11: KRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

12: KRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

13: KRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

14: KRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

15: KRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

16: KRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

1  
AA005081 ck: 2260 len: 40 ! Aa005081 Human polypeptide SPQ ID NO 18  
(R,K){20}

9: SRASP KKKKKKKKKKKKKKKKKKK  
(K){20}

10: RASPK KKKKKKKKKKKKKKKKKKK  
(K){20}

11: ASPKK KKKKKKKKKKKKKKKKKKK  
(K){20}

12: SPKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

13: PKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

14: KRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

15: KRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

16: KRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

17: KRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

18: KRSKK KKKKKKKKKKKKKKKKKKK  
(K){20}

1  
AA005130 ck: 6597 len: 116 ! Aa005130 Human polypeptide SEQ ID NO 19  
(R,K){20}

14: ETPFQ KKKKKKKKKKKKKKKKKKK  
(K){20}

15: TPFRK KKKKKKKKKKKKKKKKKKK  
(K){20}

16: PFRKK KKKKKKKKKKKKKKKKKKK  
(K){20}

1  
AA005191 ck: 4399 len: 85 ! Aa005191 Human polypeptide SPQ ID NO 19  
(R,K){20}

30: GYTQL KKKKKKKKKKKKKKKKKKK  
(R,K){20}

1  
AA005260 ck: 3618 len: 55 ! Aa005260 Human polypeptide SEQ ID NO 19  
(R,K){20}

15: PHKQ KKKKKKKKKKKKKKKKKKK  
(K){20}

(K){20}

```
16: EKHOK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
17: KHOKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
18: HOKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
19: OKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKKK KKGKG
    (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKKK KGGGD
    (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKKK GGSGD
    (K){20}

AAO05369 ck: 1417 len: 70 1 Aao05369 Human polypeptide SEQ ID NO 19261.
    (R,K){20}
22: KYHET KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
23: YHETK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
24: HETKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
25: ETKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
26: TKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKKK KKGGE
    (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKKK KKGKK
    (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKKK KGKKK
    (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKKK GKKKK
    (R,K){20}

AAO05372 ck: 9880 len: 43 1 Aao05372 Human polypeptide SEQ ID NO 19264.
    (R,K){20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
2: K KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
```

```
5: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
11: OKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}

AAO05384 ck: 715 len: 23 1 Aao05384 Human polypeptide SEQ ID NO 192
    (R,K){20}
4: LFS KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}

AAO05499 ck: 7813 len: 76 1 Aao05499 Human polypeptide SEQ ID NO 193
    (R,K){20}
7: VOPOQ KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
8: OPOOK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
9: POOKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
10: OOKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
11: OKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
```

15: KKKK (K){20} KKKK  
16: KKKK (K){20} KKKK  
17: KKKK (K){20} KKKK  
18: KKKK (K){20} KKKK  
19: KKKK (K){20} KKKK  
20: KKKK (K){20} KKKK  
21: KKKK (K){20} KKKK  
22: KKKK (K){20} KKTAS  
23: KKKK (K){20} KKASP  
24: KKKK (K){20} KASPG  
25: KKKK (K){20} ASPGG  
AA05500 ck: 7443 len: 53 ! Aa05500 Human polypeptide SEQ ID NO 19392.  
1 (R,K){20}  
5: QPOS (K){20} KKASQ  
6: QPOS (K){20} KASQK  
7: POSK (K){20} ASQKK  
AA05530 ck: 7354 len: 75 ! Aa05530 Human polypeptide SEQ ID NO 19422.  
1 (R,K){20}  
11: TTYFL (K){20} KKKK  
12: TYFLK (K){20} KKKK  
13: YFLK (K){20} KKKK  
14: FLKK (K){20} KKKK  
15: LKKK (K){20} KKKK  
16: KKKK (K){20} KKKK  
17: KKKK (K){20} KKKK

18: KKKK KKKK KKKK KKKK KKKK  
19: KKKK (K){20} KKKK  
20: KKKK (K){20} KKKK  
21: KKKK (K){20} KKKK  
22: KKKK (K){20} KKKK  
23: KKKK (K){20} KKKK  
24: KKKK (K){20} KKKK  
25: KKKK (K){20} KKKK  
26: KKKK (K){20} KKKK  
27: KKKK (K){20} KKKK  
28: KKKK (K){20} KKKK  
29: KKKK (K){20} KKKK  
30: KKKK (K){20} KKKK  
31: KKKK (K){20} KKKK  
32: KKKK (K){20} KKKK  
33: KKKK (K){20} KKKK  
34: KKKK (K){20} KKKK  
35: KKKK (K){20} KKKK  
36: KKKK (K){20} KKKK  
37: KKKK (K){20} KKKK  
38: KKKK (K){20} KKKK  
39: KKKK (K){20} KKKK  
40: KKKK (K){20} KKTAS  
41: KKKK (K){20} KASPP  
42: KKKK (K){20} ASPGG



100: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
101: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
102: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
103: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
104: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
105: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
106: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
107: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

AA007505 ck: 3528 len: 72 1 Aa007505 Human polypeptide SEQ ID NO 21397.

1  
(R,K){20}  
14: HLVEA KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: LVEAK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
16: VEAKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: EAKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: AKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

AA007509 ck: 4998 len: 140 1 Aa007509 Human polypeptide SEQ ID NO 21401.

1  
(R,K){20}  
17: KVXKE KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: VKXKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
19: KXKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
20: XKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

(K){20}  
21: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
27: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
28: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
29: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
30: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
31: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
32: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

AA007594 ck: 7271 len: 93 1 Aa007594 Human polypeptide SEQ ID NO 21.

1  
(R,K){20}  
9: FLLGG KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: LLLGG KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: LLGGK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
12: LGKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
13: GKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

AA007607 ck: 2434 len: 72 1 Aa007607 Human polypeptide SEQ ID NO 21.

1  
(R,K){20}  
21: KKKKE KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
48: GLFXE KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

AA007610 ck: 5270 len: 74 1 Aa007610 Human polypeptide SEQ ID NO 21.

1  
(R,K){20}  
19: TELTI KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

```
20: ELTIK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
21: LTIKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
22: TIKKK KKKKKKKKKKKKKKKKKKK KKKKG
      (K){20}
23: IKKKK KKKKKKKKKKKKKKKKKKK KKKGG
      (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKGCG
      (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKK KGGGA
      (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKK GGGAL
      (K){20}
AAO07755 ck: 6069 len: 90 i Aao07755 Human polypeptide SEQ ID NO 21647.
      (R,K){20}
35: KNOSS KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
36: NQSSK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
37: QSSKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
38: SSKKK KKKKKKKKKKKKKKKKKKK KKKKG
      (K){20}
39: SKKKK KKKKKKKKKKKKKKKKKKK KKKGG
      (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKK KKGGP
      (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKK KGGPL
      (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKK GGPLL
      (K){20}
AAO07762 ck: 5115 len: 41 i Aao07762 Human polypeptide SEQ ID NO 21654.
      (R,K){20}
18: IPSLX KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
19: PSLXX KKKKKKKKKKKKKKKKKKK KKK
      (K){20}
20: SLXXX KKKKKKKKKKKKKKKKKKK KK
      (K){20}
21: LKKKK KKKKKKKKKKKKKKKKKKK K
      (K){20}
22: XXXKK KKKKKKKKKKKKKKKKKKK
      (K){20}
AAO07775 ck: 9827 len: 90 i Aao07775 Human polypeptide SEQ ID NO 21667.
      (R,K){20}
46: PHXPE KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
```

```
47: HXPEK KKKKKKKKKKKKKKKKKKK KKKKG
      (K){20}
48: XPEKK KKKKKKKKKKKKKKKKKKK KKKGG
      (K){20}
49: PEKKK KKKKKKKKKKKKKKKKKKK KKGGA
      (K){20}
50: EKKKK KKKKKKKKKKKKKKKKKKK KGGAF
      (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKK GGAFF
      (K){20}
AAO07806 ck: 1959 len: 99 i Aao07806 Human polypeptide SEQ ID NO 216
      (R,K){20}
41: KLHGI KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
42: LHCIK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
43: HCIKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
44: CIKKK KKKKKKKKKKKKKKKKKKK KKKKL
      (K){20}
45: IKKKK KKKKKKKKKKKKKKKKKKK KKKLW
      (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKK KKLWG
      (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKK KLWGG
      (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKK LMGGG
      (K){20}
AAO07872 ck: 8003 len: 135 i Aao07872 Human polypeptide SEQ ID NO 217
      (R,K){20}
52: TKKKQ KKKKKKKKKKKKKKKKKKK GGGLL
      (K){20}
AAO07874 ck: 8007 len: 111 i Aao07874 Human polypeptide SEQ ID NO 217
      (R,K){20}
90: EEPST KKKKKKKKKKKKKKKKKKK KG
      (K){20}
91: EPSIK KKKKKKKKKKKKKKKKKKK G
      (K){20}
AAO07876 ck: 8130 len: 66 i Aao07876 Human polypeptide SEQ ID NO 217
      (R,K){20}
17: IYSPF KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
18: YSPEK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
19: SPEKK KKKKKKKKKKKKKKKKKKK KKKKG
      (K){20}
```



20: FEKKK (K){20} KKKKKKKKKKKKKKKKK KKKOK  
 21: EKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
 22: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKE  
 23: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK OKKEN

AA007976 ck: 6471 len: 97 ! Aa007976 Human polypeptide SEQ ID NO 21868.

(R,K){20}

43: PPIST (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

44: PISTK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

45: ISTKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

46: STKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

47: TKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

50: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

51: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

52: KKKKK KKKKKKKKKKKKKKKKKKK KKKAS

(K){20}

53: KKKKK KKKKKKKKKKKKKKKKKKK KKKAS

(K){20}

54: KKKKK KKKKKKKKKKKKKKKKKKK KASSG

(K){20}

55: KKKKK KKKKKKKKKKKKKKKKKKK ASSGG

AA007981 ck: 9551 len: 81 ! Aa007981 Human polypeptide SEQ ID NO 21873.

(R,K){20}

54: KPCIY KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

55: PCIYK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

56: CIYKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

57: IYKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

58: YKKKK KKKKKKKKKKKKKKKKKKK KKKK  
 59: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKK  
 60: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KK  
 61: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK K  
 62: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK

AA008018 ck: 9768 len: 48 ! Aa008018 Human polypeptide SEQ ID NO 219

(R,K){20}

24: KISQV (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

25: ISOVR KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

26: SQVRK KKKKKKKKKKKKKKKKKKK KKK

(K){20}

27: QVRKK KKKKKKKKKKKKKKKKKKK KK

(K){20}

28: VRKKK KKKKKKKKKKKKKKKKKKK K

(K){20}

29: RKKKK KKKKKKKKKKKKKKKKKKK

AA008121 ck: 5381 len: 86 ! Aa008121 Human polypeptide SEQ ID NO 220

(R,K){20}

20: HFGLL (K){20} KKKKKKKKKKKKKKKKKKK GGGA

AA008232 ck: 782 len: 104 ! Aa008232 Human polypeptide SEQ ID NO 221

(R,K){20}

49: ISQON KKKKKKKKKKKKKKKKKKK KDSGG

(K){20}

50: SQQNK KKKKKKKKKKKKKKKKKKK DGGGA

AA008450 ck: 6228 len: 31 ! Aa008450 Human polypeptide SEQ ID NO 223

(R,K){20}

5: NNSA KKKKKKKKKKKKKKKKKKK KKGGA

(K){20}

6: NNSAK KKKKKKKKKKKKKKKKKKK KGGAA

(K){20}

7: NSAKK KKKKKKKKKKKKKKKKKKK GGAAF

AA008458 ck: 1813 len: 59 ! Aa008458 Human polypeptide SEQ ID NO 223

(R,K){20}

8: LGWE KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}  
9: GWEK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
10: GWEK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
11: WEKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
12: EKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
13: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
AA008460 ck: 5334 len: 62 1 Aa008460 Human polypeptide SEQ ID NO 22352.  
(R,K){20}  
22: KEFT KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
23: EFST KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
24: FSTK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
25: STKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
26: TKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
27: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
28: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
29: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
30: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
31: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
32: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
33: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
34: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
AA008466 ck: 8728 len: 107 1 Aa008466 Human polypeptide SEQ ID NO 22358.  
(R,K){20}  
57: FWGV KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
58: WGVK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
59: GGVK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
60: GVKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}  
61: VKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
62: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(R,K){20}  
63: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(R,K){20}  
64: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(R,K){20}  
65: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
AA008580 ck: 5164 len: 62 1 Aa008580 Human polypeptide SEQ ID NO 224  
(R,K){20}  
30: PILL KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
AA008591 ck: 900 len: 46 1 Aa008591 Human polypeptide SEQ ID NO 224  
(R,K){20}  
14: LLCS KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
AA008609 ck: 5151 len: 74 1 Aa008609 Human polypeptide SEQ ID NO 225  
(R,K){20}  
19: KCVL KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
19: CVLK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
AA008612 ck: 557 len: 46 1 Aa008612 Human polypeptide SEQ ID NO 225  
(R,K){20}  
20: TFCM KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
21: FCIM KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
22: CIMK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
23: IMKK KKKKKKKKKKKKKKKKKKKKKKKKK  
AA008616 ck: 4555 len: 41 1 Aa008616 Human polypeptide SEQ ID NO 225  
(R,K){20}  
6: HCALP KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
7: CALP KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
8: ALPK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
9: LPKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

10: PKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
19: KKKKK KKKKKKKKKKKKKKKKKKK KKK  
(K){20}  
20: KKKKK KKKKKKKKKKKKKKKKKKK KK  
(K){20}  
21: KKKKK KKKKKKKKKKKKKKKKKKK K  
(K){20}  
22: KKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

AAO08623 ck: 9243 len: 119 1 Aao08623 Human polypeptide SEQ ID NO 22515.  
(R,K){20}  
(R,K){20}

21: DSKQZ KKKKKKKKKKKKKKKKKKKR GGAFK  
(R,K){20}

AAO08624 ck: 5590 len: 65 1 Aao08624 Human polypeptide SEQ ID NO 22516.  
(R,K){20}  
(K){20}

11: EGNMW KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

12: GNNWK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

13: NNWKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

14: NNWKK KKKKKKKKKKKKKKKKKKK KKWGG  
(K){20}

15: WKKKK KKKKKKKKKKKKKKKKKKK KWWGA  
(K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKK WGGAL  
(K){20}

AAO08631 ck: 6430 len: 135 1 Aao08631 Human polypeptide SEQ ID NO 22523.  
(R,K){20}  
(K){20}

30: TPSRA KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

31: PSRAK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

32: SRAKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

33: RAKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

34: AKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

35: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

36: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

37: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

38: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

39: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

40: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

41: KKKKK KKKKKKKKKKKKKKKKKKKKK KRRGG  
(K){20}

42: KKKKK KKKKKKKKKKKKKKKKKKKKK KRRGG  
(K){20}

43: KKKKK KKKKKKKKKKKKKKKKKKKKK RGGGP  
(R,K){20}

44: KKKKK KKKKKKKKKKKKKKKKKKKKKR GGSPK  
(R,K){20}

AAO08644 ck: 3202 len: 80 1 Aao08644 Human polypeptide SEQ ID NO 225  
(R,K){20}  
(K){20}

2: E KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

3: EK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

4: EKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

5: EKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

6: EKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

11: KKKK (K){20} KKKK  
 12: KKKK (K){20} KKKK  
 13: KKKK (K){20} KKKK  
 14: KKKK (K){20} KKKK  
 15: KKKK (K){20} KKKK  
 16: KKKK (K){20} KKKG  
 17: KKKK (K){20} KKGK  
 18: KKKK (K){20} KGGG  
 19: KKKK (K){20} GGGP  
 AAO08653 ck: 6448 len: 63 1 Aao08653 Human polypeptide SEQ ID NO 22545  
 12: PIRLT (R,K){20} KKKK  
 13: IKLTK (K){20} KKKK  
 14: KLTKK (K){20} KKKK  
 15: LTKKK (K){20} KKKK  
 16: TKKKK (K){20} KKKK  
 17: KKKKK (K){20} KKKK  
 18: KKKKK (K){20} KKKK  
 19: KKKKK (K){20} KKKK

1  
 AAO08686 ck: 6936 len: 81 1 Aao08686 Human polypeptide SEQ ID NO 22578  
 23: HPVNH (R,K){20} KKKK  
 24: PVNHK (K){20} KKKK  
 25: VNHKK (K){20} KKKK

26: NHKK (K){20} KKKK  
 27: HKKK (K){20} KKKK  
 28: KKKK (K){20} KKKK  
 29: KKKK (K){20} KKKK  
 30: KKKK (K){20} KKKK  
 31: KKKK (K){20} KKKK  
 32: KKKK (K){20} KKKK  
 33: KKKK (K){20} KKKK  
 34: KKKK (K){20} KKKK  
 35: KKKK (K){20} KKEGA  
 36: KKKK (K){20} KEGAL  
 37: KKKK (K){20} EGALG  
 AAO08701 ck: 588 len: 46 1 Aao08701 Human polypeptide SEQ ID NO 225

1  
 7: INANS (R,K){20} KKKK  
 8: NANSK (K){20} KKKK  
 9: ANSKK (K){20} KKKK  
 10: NSKKK (K){20} KKKK  
 11: SKKKK (K){20} KKGK  
 12: KKKKK (K){20} KGGK  
 13: KKKKK (K){20} GGGK  
 AAO08707 ck: 9359 len: 48 1 Aao08707 Human polypeptide SEQ ID NO 225  
 13: TAVTQ (R,K){20} KKKK  
 14: AVTQK (K){20} KKGK  
 15: VTQKK (K){20} GGGK

1  
AAO08708 ck: 9654 len: 32 ! Aao08708 Human polypeptide SEQ ID NO 22600.  
(R,K){20}  
(K){20}  
8: SMTEP KKKKKKKKKKKKKKKKKKK  
(K){20}  
9: MTEPK KKKKKKKKKKKKKKKKKKK  
(K){20}  
10: TEPK KKKKKKKKKKKKKKKKKKK  
(K){20}  
11: EPKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
12: PKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
13: KKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
1  
AAO08717 ck: 4688 len: 99 ! Aao08717 Human polypeptide SEQ ID NO 22609.  
(R,K){20}  
(K){20}  
80: SPILLI KKKKKKKKKKKKKKKKKKK  
(K){20}  
1  
AAO08772 ck: 1431 len: 105 ! Aao08772 Human polypeptide SEQ ID NO 22664.  
(R,K){20}  
(K){20}  
23: IMNYL KKKKKKKKKKKKKKKKKKK  
(K){20}  
24: MNYLK KKKKKKKKKKKKKKKKKKK  
(K){20}  
25: NYLKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
26: YLKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
27: LKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
28: KKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
29: KKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
30: KKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
1  
AAO08774 ck: 2529 len: 71 ! Aao08774 Human polypeptide SEQ ID NO 22666.  
(R,K){20}  
(K){20}  
16: RTIKL KKKKKKKKKKKKKKKKKKK  
(K){20}  
17: TIKLK KKKKKKKKKKKKKKKKKKK  
(K){20}  
18: IKLKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
19: KLKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

1  
20: LKKKK (K){20} KKKKKKKKKKKKKKKKKKK  
(K){20}  
21: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK  
(K){20}  
22: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK  
(K){20}  
23: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK  
(K){20}  
1  
AAO08817 ck: 529 len: 28 ! Aao08817 Human polypeptide SEQ ID NO 227.  
(R,K){20}  
(K){20}  
3: AX KKKKKKKKKKKKKKKKKKK  
(K){20}  
1  
AAO08820 ck: 1888 len: 84 ! Aao08820 Human polypeptide SEQ ID NO 227.  
(R,K){20}  
(K){20}  
28: CMTPS KKKKKKKKKKKKKKKKKKK  
(K){20}  
29: MTFKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
1  
AAO08841 ck: 3529 len: 74 ! Aao08841 Human polypeptide SEQ ID NO 227.  
(R,K){20}  
(K){20}  
3: PQ KKKKKKKKKKKKKKKKKKK  
(K){20}  
4: POK KKKKKKKKKKKKKKKKKKK  
(K){20}  
1  
AAO08857 ck: 2816 len: 49 ! Aao08857 Human polypeptide SEQ ID NO 227.  
(R,K){20}  
(K){20}  
25: PPPTS KKKKKKKKKKKKKKKKKKK  
(K){20}  
26: PPTSK KKKKKKKKKKKKKKKKKKK  
(K){20}  
27: PLSKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
28: TSKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
29: SKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
30: KKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
1  
AAO08913 ck: 3132 len: 52 ! Aao08913 Human polypeptide SEQ ID NO 228.  
(R,K){20}  
(K){20}  
9: IIKSF KKKKKKKKKKKKKKKKKKK  
(K){20}  
10: IKSF KKKKKKKKKKKKKKKKKKK  
(K){20}  
11: KSFKK KKKKKKKKKKKKKKKKKKK  
(K){20}

12: SFRKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
13: FRKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
14: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
15: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
16: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
17: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
18: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
19: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
20: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
21: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
22: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
23: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
24: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
25: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
26: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK  
27: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKI  
28: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKI  
29: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKII  
30: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KII  
31: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK II

1 AAO08914 ck: 4641 len: 25 ! Aao08914 Human polypeptide SEQ ID NO 22806  
(R,K){20}  
(R,K){20}

6: ILMPX KKKKKKKKKKKKKKKKKKK

1 AAO08943 ck: 5770 len: 75 ! Aao08943 Human polypeptide SEQ ID NO 22835  
(R,K){20}  
(K){20}

8: SEMAA KKKKKKKKKKKKKKKKKKK KKKIF  
9: EWAAR (K){20} KKKKKKKKKKKKKKKKKKK KKIFP  
10: WAAKK (K){20} KKKKKKKKKKKKKKKKKKK KIFFP  
11: AAARR (K){20} KKKKKKKKKKKKKKKKKKK IFFPG

1 AAO08994 ck: 2695 len: 29 ! Aao08994 Human polypeptide SEQ ID NO 228  
(R,K){20}  
(K){20}

9: NLPSS KKKKKKKKKKKKKKKKKKK K  
10: LPSSK (K){20} KKKKKKKKKKKKKKKKKKK

1 AAO08995 ck: 6270 len: 26 ! Aao08995 Human polypeptide SEQ ID NO 228  
(R,K){20}  
(K){20}

6: MAAPP KKKKKKKKKKKKKKKKKKK I

1 AAO09001 ck: 6400 len: 26 ! Aao09001 Human polypeptide SEQ ID NO 228  
(R,K){20}  
(K){20}

5: LTSS KKKKKKKKKKKKKKKKKKK KK  
6: LTSSK (K){20} KKKKKKKKKKKKKKKKKKK K  
7: TSSKK (K){20} KKKKKKKKKKKKKKKKKKK

1 AAO09016 ck: 4378 len: 119 ! Aao09016 Human polypeptide SEQ ID NO 229  
(R,K){20}  
(R,K){20}

10: EEEEE RRRRRRRRRRRRRRRRRR ILRQK

1 AAO09057 ck: 3712 len: 58 ! Aao09057 Human polypeptide SEQ ID NO 229  
(R,K){20}  
(K){20}

5: MILN KKKKKKKKKKKKKKKKKKK KKKKK  
6: MILNK (K){20} KKKKKKKKKKKKKKKKKKK KKKKG  
7: ILNKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKG  
8: LNKRR (K){20} KKKKKKKKKKKKKKKKKKK KKGKG  
9: NKKKK (K){20} KKKKKKKKKKKKKKKKKKK KGGGP  
10: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK GGKGP

1 AAO09066 ck: 2645 len: 29 ! Aao09066 Human polypeptide SEQ ID NO 229

```
1
  4: AQQ KKKKKKKKKKKKKKKKKKKKKKKKK
      (R,K){20}
      (K){20}
  5: AQQ KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
  6: AQQK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
  7: QQQK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
  8: QKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
  9: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 10: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
AA009072 ck: 8432 len: 42 1 Aa009072 Human polypeptide SEQ ID NO 22964.
      (R,K){20}
      (K){20}
 11: LIPTL KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 12: IFTLK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 13: FTLLK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 14: TLKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
AA009077 ck: 9524 len: 87 1 Aa009077 Human polypeptide SEQ ID NO 22969.
      (R,K){20}
      (K){20}
 26: RFLTL KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 27: FLTLK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 28: LLTK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 29: LTKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 30: TKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
AA009162 ck: 3841 len: 100 1 Aa009162 Human polypeptide SEQ ID NO 23054.
      (R,K){20}
      (K){20}
 11: KTLFQ KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 12: TLFQK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 13: LFQKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 14: FQKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
```

```
1
 15: QKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 16: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 17: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 18: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 19: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 20: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 21: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 22: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 23: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 24: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 25: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 26: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 27: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 28: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 29: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (R,K){20}
AA009258 ck: 7988 len: 42 1 Aa009258 Human polypeptide SEQ ID NO 23:
      (R,K){20}
      (K){20}
  7: LFCP KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
  8: LFCPK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
  9: FCPK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 10: CFPK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 11: PKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
 12: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
AA009269 ck: 7553 len: 106 1 Aa009269 Human polypeptide SEQ ID NO 23:
      (R,K){20}
      (R,K){20}
 39: NTGML KKKKKKKKKKKKKKKKKKKKKKKR
      (R,K){20}
```

1 AAC09457 ck: 9658 len: 32 | Aao09457 Human polypeptide SEQ ID NO 23349.  
(R,K){20}  
(K){20}  
9: ALVPQ KKKKKKKKKKKKKKKKKKK NIKI

1 AAC09819 ck: 4709 len: 80 | Aao09819 Human polypeptide SEQ ID NO 23711.  
(R,K){20}  
(K){20}  
54: CEFVX KKKKKKKKKKKKKKKKKKK EKGGS

1 AA010447 ck: 6342 len: 50 | Aao10447 Human polypeptide SEQ ID NO 24339.  
(R,K){20}  
(K){20}  
25: RFLKL KKKKKKKKKKKKKKKKKKK RKKKK  
(R,K){20}  
26: FLKLL KKKKKKKKKKKKKKKKKKK KKKKI  
(R,K){20}  
27: LKLKK KKKKKKKKKKKKKKKKKKK KKKI  
(R,K){20}  
28: KLKKK KKKKKKKKKKKKKKKKKKK KKI  
(R,K){20}  
29: LKKKK KKKKKKKKKKKKKKKKKKK KI  
(R,K){20}  
30: KKKKK KKKKKKKKKKKKKKKKKKK I

1 AA010451 ck: 3955 len: 82 | Aao10451 Human polypeptide SEQ ID NO 24343.  
(R,K){20}  
(K){20}  
59: SRASP KKKKKKKKKKKKKKKKKKK ARGG

1 AA010467 ck: 7542 len: 116 | Aao10467 Human polypeptide SEQ ID NO 24359.  
(R,K){20}  
(K){20}  
57: KCEFM KKKKKKKKKKKKKKKKKKK KIGGG  
(K){20}  
58: CEFMK KKKKKKKKKKKKKKKKKKK IGGGA

1 AA010564 ck: 9156 len: 98 | Aao10564 Human polypeptide SEQ ID NO 24456.  
(R,K){20}  
(K){20}  
18: NLITL KKKKKKKKKKKKKKKKKKK KKEMP  
(R,K){20}  
19: LITLK KKKKKKKKKKKKKKKKKKK KEMPV  
(K){20}  
20: LITLK KKKKKKKKKKKKKKKKKKK EMPVK

1 AA010608 ck: 9106 len: 67 | Aao10608 Human polypeptide SEQ ID NO 24500.  
(R,K){20}  
(K){20}  
42: CRUSE KKKKKKKKKKKKKKKKKKK KKKKD

1 43: RLSEK KKKKKKKKKKKKKKKKKKK KKKDS  
(K){20}  
44: LSEKK KKKKKKKKKKKKKKKKKKK KKDS  
(K){20}  
45: SEKKK KKKKKKKKKKKKKKKKKKK KDS  
(K){20}  
46: EKKKK KKKKKKKKKKKKKKKKKKK DS

1 AA010638 ck: 621 len: 56 | Aao10638 Human polypeptide SEQ ID NO 245  
(R,K){20}  
(K){20}  
33: CEPQP KKKKKKKKKKKKKKKKKKK KARG  
(K){20}  
34: EPQPK KKKKKKKKKKKKKKKKKKK ARG

1 AA010786 ck: 7349 len: 65 | Aao10786 Human polypeptide SEQ ID NO 246  
(R,K){20}  
(K){20}  
24: CPKFS KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
25: PKFSK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
26: XFSKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
27: FSKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
28: SKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
29: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
30: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
31: KKKKK KKKKKKKKKKKKKKKKKKK KKKKV  
(K){20}  
32: KKKKK KKKKKKKKKKKKKKKKKKK KKKVF  
(K){20}  
33: KKKKK KKKKKKKKKKKKKKKKKKK KKVFF  
(K){20}  
34: KKKKK KKKKKKKKKKKKKKKKKKK KVFVF  
(K){20}  
35: KKKKK KKKKKKKKKKKKKKKKKKK VVFEE

1 AA010852 ck: 5147 len: 41 | Aao10852 Human polypeptide SEQ ID NO 247  
(R,K){20}  
(K){20}  
20: FSRKM KKKKKKKKKKKKKKKKKKK KK  
(K){20}  
21: SRMKK KKKKKKKKKKKKKKKKKKK K



22: RYMKR (K){20}  
XXXXXXXXXXXXXXXXXXXX

AA010853 ck: 444 len: 98 ! Aa010853 Human polypeptide SEQ ID NO 24745.

14: LRAHL (R,K){20}  
(K){20}  
XXXXXXXXXXXXXXXXXXXX

15: RAHLK (K){20}  
XXXXXXXXXXXXXXXXXXXX

16: AHLK (K){20}  
XXXXXXXXXXXXXXXXXXXX

17: HLKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

18: LKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

19: KKKK (R){20}  
XXXXXXXXXXXXXXXXXXXX

20: KKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

21: KKKK (R){20}  
XXXXXXXXXXXXXXXXXXXX

22: KKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

23: KKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

24: KKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

25: KKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

26: KKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

27: KKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

28: KKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

29: KKKK (R,K){20}  
XXXXXXXXXXXXXXXXXXXX

AA010859 ck: 2690 len: 70 ! Aa010859 Human polypeptide SEQ ID NO 24751.

44: XGITE (R,K){20}  
(K){20}  
XXXXXXXXXXXXXXXXXXXX

45: GITEK (K){20}  
XXXXXXXXXXXXXXXXXXXX

AA010933 ck: 849 len: 69 ! Aa010933 Human polypeptide SEQ ID NO 24825.

33: FINTE (R,K){20}  
(K){20}  
XXXXXXXXXXXXXXXXXXXX

34: INTEK (K){20}  
XXXXXXXXXXXXXXXXXXXX

35: NTEKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

36: TEKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

37: EKKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

38: KKKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

39: KKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

40: KKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

41: KKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

42: KKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

43: KKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

44: KKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

45: KKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

AA010973 ck: 824 len: 51 ! Aa010973 Human polypeptide SEQ ID NO 248

16: SYFYM (R,K){20}  
(R,K){20}  
XXXXXXXXXXXXXXXXXXXX

17: YFYM (K){20}  
XXXXXXXXXXXXXXXXXXXX

AA010983 ck: 694 len: 51 ! Aa010983 Human polypeptide SEQ ID NO 248

24: HPILX (R,K){20}  
(K){20}  
XXXXXXXXXXXXXXXXXXXX

25: PILX (K){20}  
XXXXXXXXXXXXXXXXXXXX

26: ILXK (K){20}  
XXXXXXXXXXXXXXXXXXXX

27: LKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

28: XKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

29: KKKK (K){20}  
XXXXXXXXXXXXXXXXXXXX

AA010995 ck: 2382 len: 122 ! Aa010995 Human polypeptide SEQ ID NO 248

22: PCKGS (R,K){20}  
(K){20}  
XXXXXXXXXXXXXXXXXXXX

(K){20}  
23: CKGSK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
24: KGSKK KKKKKKKKKKKKKKKKK KKKKG  
(K){20}  
25: GSKKK KKKKKKKKKKKKKKKKK KKKKG  
(K){20}  
26: SKKKK KKKKKKKKKKKKKKKKK KKGKG  
(K){20}  
27: KKKKK KKKKKKKKKKKKKKKKK KGGGG  
(K){20}  
28: KKKKK KKKKKKKKKKKKKKKKK GGGGF

AA01097 ck: 5315 len: 58 1 Aa01097 Human polypeptide SEQ ID NO 24899.  
(R,K){20}  
29: IYIIQ KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
30: XIIOK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
31: IIOKK KKKKKKKKKKKKKKKKK KKKKG  
(K){20}  
32: IOKKK KKKKKKKKKKKKKKKKK KKRGG  
(K){20}  
33: QKKKK KKKKKKKKKKKKKKKKK KRGGP  
(K){20}  
34: KKKKK KKKKKKKKKKKKKKKKK RGPPP  
(R,K){20}  
35: KKKKK KKKKKKKKKKKKKKKKK GGPPP

AA01102 ck: 8638 len: 82 1 Aa01102 Human polypeptide SEQ ID NO 24894.  
(R,K){20}  
5: FKRS KKKKKKKKKKKKKKKKKR GGALL  
(R,K){20}  
38: XFRRS KKKKKKKKKKKKKKKKKR GGALL

AA01103 ck: 3780 len: 49 1 Aa01103 Human polypeptide SEQ ID NO 24925.  
(R,K){20}  
10: QKKFI RKKKKKKKKKKKKKKKKK KLSKK  
(R,K){20}  
11: XKFIR KKKKKKKKKKKKKKKKKK LSKKI  
(K){20}

AA01104 ck: 7868 len: 85 1 Aa01104 Human polypeptide SEQ ID NO 24940.  
(R,K){20}  
53: IISDP KKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
54: ISDPK KKKKKKKKKKKKKKKKK KKKKK

1

(K){20}  
55: SDPKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
56: DPKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
57: PKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
58: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
59: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
60: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
61: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
62: KKKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
63: KKKKK KKKKKKKKKKKKKKKKK KKK  
(K){20}  
64: KKKKK KKKKKKKKKKKKKKKKK KK  
(K){20}  
65: KKKKK KKKKKKKKKKKKKKKKK K  
(K){20}  
66: KKKKK KKKKKKKKKKKKKKKKK

AA01107 ck: 4186 len: 100 1 Aa01107 Human polypeptide SEQ ID NO 249  
(R,K){20}  
39: PPRPD KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
40: PRPDK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
41: KPDKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
42: PDKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
43: DKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
44: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
45: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
46: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
47: KKKKK KKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
48: KKKKK KKKKKKKKKKKKKKKKK KKKKG  
(K){20}  
49: KKKKK KKKKKKKKKKKKKKKKK KKGKG

50: KKKKK (K){20} KRGGG  
51: KKKKK (K){20} RGGGG  
52: KKKKK (R,K){20} GGGGG

AA01124 ck: 5121 len: 61 ! Aa01124 Human polypeptide SEQ ID NO 25016.

1

16: KYSFL (R,K){20} KKKKK  
(K){20}

17: YSFLK (K){20} KKKKK

18: SFLKK (K){20} KKKKK

19: FLKKK (K){20} KKKKK

20: LKKKK (K){20} KKKKK

21: KKKKK (K){20} KKKKK

22: KKKKK (K){20} KKKKK

23: KKKKK (K){20} KKKKK

24: KKKKK (K){20} KKKKK

25: KKKKK (K){20} KKKKK

26: KKKKK (K){20} KKKKK

27: KKKKK (K){20} KKKKK

28: KKKKK (K){20} KKKKK

29: KKKKK (K){20} KKKKK

30: KKKKK (K){20} KKKKK

31: KKKKK (K){20} KKKKK

32: KKKKK (K){20} KKKKK

33: KKKKK (K){20} KKKKK

34: KKKKK (K){20} KKKKK

35: KKKKK (K){20} KKKKK

36: KKKKK (K){20} KKKKK

37: KKKKK (K){20} KKKKK

38: KKKKK (K){20} KKKKK

39: KKKKK (K){20} KKKKK

40: KKKKK (K){20} KKKKK

41: KKKKK (K){20} KKKKK

42: KKKKK (K){20} KKKKK

AA01139 ck: 8807 len: 68 ! Aa01139 Human polypeptide SEQ ID NO 250

1

22: VCHFP (R,K){20} KKKKK  
(K){20}

23: CREPK (K){20} KKKKK

AA01165 ck: 7203 len: 62 ! Aa01165 Human polypeptide SEQ ID NO 250

1

18: DASMV (R,K){20} KKKKK  
(K){20}

19: ASMVK (K){20} KKKKK

20: SMVKK (K){20} KKKKK

21: MVKKK (K){20} KKKKK

22: VKKKK (K){20} KKKKK

23: KKKKK (K){20} KKKKK

24: KKKKK (K){20} KKKKK

25: KKKKK (K){20} KKKKK

26: KKKKK (K){20} KKKKK

AA01109 ck: 6562 len: 96 ! Aa01109 Human polypeptide SEQ ID NO 251

1

63: LLAYK (R,K){20} KKKKK  
(K){20}

64: LLAYK (K){20} KKKKK

```

65: LAykk (K){20}
AAAAAAAAAAAAAAAAAAAA PkYLS

1
AA011210 ck: 863 len: 70 1 Aa011210 Human polypeptide SEQ ID NO 25102.
(R,K){20}
31: DLGLC (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
32: LCLCk (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
33: CLCkK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
34: LCKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
35: CKKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
36: KKKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
37: KKKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
38: KKKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
39: KKKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
40: KKKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
41: KKKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
42: KKKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
43: KKKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
44: KKKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
45: KKKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
AAAAAAAAAAAAAAAAAAAA PGGGG

AA011214 ck: 9584 len: 68 1 Aa011214 Human polypeptide SEQ ID NO 25106.
(R,K){20}
19: NRVCx (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
20: RVCxK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
21: VCxKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
22: CXKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
23: xKKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
24: KKKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
AAAAAAAAAAAAAAAAAAAA RGGPL
```

```

25: KKKKK (R,K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
AAAAAAAAAAAAAAAAAAAA GGPLK

1
AA011241 ck: 1100 len: 36 1 Aa011241 Human polypeptide SEQ ID NO 251
(R,K){20}
8: XQTLT (K){20}
AAAAAAAAAAAAAAAAAAAA KAYTS
(K){20}
9: QTLTK (K){20}
AAAAAAAAAAAAAAAAAAAA AYTSK

1
AA011248 ck: 5013 len: 88 1 Aa011248 Human polypeptide SEQ ID NO 251
(R,K){20}
51: TCLYx (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
52: CLYxK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
53: LYxKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
54: YxKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
55: xKKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
56: KKKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
57: KKKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
58: KKKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
(K){20}
59: KKKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKKK
AAAAAAAAAAAAAAAAAAAA GGAFF

1
AA011293 ck: 6490 len: 73 1 Aa011293 Human polypeptide SEQ ID NO 251
(R,K){20}
46: KXDYF (K){20}
AAAAAAAAAAAAAAAAAAAA KKKSG
(K){20}
47: XDYFK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKSG
(K){20}
48: DYFKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKSG
(K){20}
49: YFKKK (K){20}
AAAAAAAAAAAAAAAAAAAA KKKSG

1
AA011324 ck: 6024 len: 93 1 Aa011324 Human polypeptide SEQ ID NO 252
(R,K){20}
37: ISLSA (R,K){20}
AAAAAAAAAAAAAAAAAAAA KKKGG
(K){20}
38: SLSAR (K){20}
AAAAAAAAAAAAAAAAAAAA KKKGG
(K){20}
39: LSARx (K){20}
AAAAAAAAAAAAAAAAAAAA KKKGG
```

40: SARKK (K){20}  
KKKKKKKKKKKKKKKKKK GGGGV

AA011327 ck: 3052 len: 102 ! Aa011327 Human polypeptide SEQ ID NO 25219.

(R,K){20}

20: LSKEL (K){20}  
KKKKKKKKKKKKKKKKKK KKKLG

(K){20}

21: SKELK KKKKKKKKKKKKKKKKK KKLGD

(K){20}

22: KELKK KKKKKKKKKKKKKKKKK KLGDE

(K){20}

23: ELKKK KKKKKKKKKKKKKKKKK LGDEE

AA011342 ck: 7189 len: 106 ! Aa011342 Human polypeptide SEQ ID NO 25234.

(R,K){20}

61: FFKTX KKKKKKKKKKKKKKKKK GGPLK

AA011346 ck: 7311 len: 85 ! Aa011346 Human polypeptide SEQ ID NO 25238.

(R,K){20}

62: ANHWE KKKKKKKKKKKKKKKKK SGGG

AA011352 ck: 1342 len: 51 ! Aa011352 Human polypeptide SEQ ID NO 25244.

(R,K){20}

25: SIFKX KKKKKKKKKKKKKKKKK KNPPS

(K){20}

26: IFKKK KKKKKKKKKKKKKKKKK NPPSL

AA011705 ck: 7660 len: 101 ! Aa011705 Human polypeptide SEQ ID NO 25597.

(R,K){20}

17: WAPLQ KKKKKKKKKKKKKKKKK KKKGG

(K){20}

18: APLQK KKKKKKKKKKKKKKKKK KKGGA

(K){20}

19: PLQKK KKKKKKKKKKKKKKKKK KGGAP

(K){20}

20: LQKKK KKKKKKKKKKKKKKKKK GGAPL

AA011799 ck: 8286 len: 39 ! Aa011799 Human polypeptide SEQ ID NO 25691.

(R,K){20}

7: NPPVS RKKKKKKKKKKKKKKKK KKKKK

(K){20}

8: PPSVR KKKKKKKKKKKKKKKKK KKKKK

(K){20}

9: PPSRK KKKKKKKKKKKKKKKKK KKKKK

10: VSRKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

11: SRKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

12: RKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKK

13: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKGG

14: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKGCG

15: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KGGGL

16: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK GGGL

AA011820 ck: 8725 len: 42 ! Aa011820 Human polypeptide SEQ ID NO 257

(R,K){20}

17: NFHAL KKKKKKKKKKKKKKKKK SSSLR

AA011828 ck: 4807 len: 41 ! Aa011828 Human polypeptide SEQ ID NO 257

(R,K){20}

1: KKKKK KKKKKKKKKKKKKKKKK KKKKK

2: K KKKKKKKKKKKKKKKKK KKKKK

3: KK KKKKKKKKKKKKKKKKK KKKKK

4: KKK KKKKKKKKKKKKKKKKK KKKKK

5: KKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKRG

6: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKRGG

7: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KRGGG

8: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK RGGGF

9: KKKKK (R,K){20}  
KKKKKKKKKKKKKKKKKKR GGGFK

AA011831 ck: 1761 len: 29 ! Aa011831 Human polypeptide SEQ ID NO 257

(R,K){20}

1: KKKKK (K){20}  
KKKKKKKKKKKKKKKKKK KKKKA

2: K (K){20}  
KKKKKKKKKKKKKKKKKK KKKAG

3: KK (K){20}  
KKKKKKKKKKKKKKKKKK KKAAG

4: KKK KKKKKKKKKKKKKKKKK KAGG  
(K){20}  
5: KKK KKKKKKKKKKKKKKKKK AGCG

AA011843 ck: 86 len: 36 1 Aa011843 Human polypeptide SEQ ID NO 25735.

1

(R,K){20}  
(K){20}  
1: KKKKKKKKKKKKKKKKK KKKK

(K){20}  
2: K KKKKKKKKKKKKKKKKK KKKK

(K){20}  
3: KK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
4: KKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
5: KKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
6: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
7: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
8: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
9: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
10: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
11: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
12: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
13: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
14: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
15: KKKK KKKKKKKKKKKKKKKKK KKKK

AA011844 ck: 3574 len: 49 1 Aa011844 Human polypeptide SEQ ID NO 25736.

1

(R,K){20}  
(K){20}  
18: LLYME KKKKKKKKKKKKKKKKK KKKK

(K){20}  
19: LYMEK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
20: YMEKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
21: MEKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
22: EKKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
23: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
24: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
25: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
26: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
27: KKKK KKKKKKKKKKKKKKKKK KKKK

AA011845 ck: 8375 len: 45 1 Aa011845 Human polypeptide SEQ ID NO 257

1

(R,K){20}  
(K){20}  
15: SPSTL KKKKKKKKKKKKKKKKK KKKK

(K){20}  
16: PSTLK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
17: STLKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
18: TLKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
19: LKKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
20: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
21: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
22: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
23: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}  
24: KKKK KKKKKKKKKKKKKKKKK KKKK

AA011849 ck: 284 len: 36 1 Aa011849 Human polypeptide SEQ ID NO 257

1

(R,K){20}  
(K){20}  
13: KCLCE KKKKKKKKKKKKKKKKK KKKK

(K){20}  
14: CLCEK KKKKKKKKKKKKKKKKK KKKK

(R,K){20}  
15: LCEKK KKKKKKKKKKKKKKKKK KKKK

(R,K){20}  
16: CEKKK KKKKKKKKKKKKKKKKK KKKK

(R,K){20}  
17: EKKKK KKKKKKKKKKKKKKKKK KKKK

AA011886 ck: 5282 len: 97 1 Aa011886 Human polypeptide SEQ ID NO 257

1

(R,K){20}  
(K){20}

```
32: SNSSL KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
33: NSSLK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
34: SSLK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
35: SLKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
36: LKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
37: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
38: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
39: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
40: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
41: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
42: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
43: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
44: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
45: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
46: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}

AA01868 ck: 6488 len: 31 1 Aa01868 Human polypeptide SEQ ID NO 25780.
      (R,K){20}
      (K){20}
2: D KKKKKKKKKKKKKKKKKKK KAGGG
      (K){20}
3: DK KKKKKKKKKKKKKKKKKKK AGGGL

AA01903 ck: 8876 len: 48 1 Aa01903 Human polypeptide SEQ ID NO 25795.
      (R,K){20}
      (K){20}
17: VIRYV KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
18: IKYVK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
19: KYVK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
20: YVKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
21: VKKK KKKKKKKKKKKKKKKKKKK KKKKK
```

```
22: KKKK (K){20}
23: KKKK (K){20}
24: KKKK (K){20}
25: KKKK (K){20}
26: KKKK (K){20}
27: KKKK (K){20}
28: KKKK (K){20}
29: KKKK (K){20}

AA01999 ck: 3954 len: 47 1 Aa01999 Human polypeptide SEQ ID NO 251
      (R,K){20}
      (K){20}
17: IDEGL KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
18: DEGLK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
19: EGLKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
20: GLKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
21: LKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
22: KKKK KKKKKKKKKKKKKKKKKKK KKGGA
      (K){20}
23: KKKK KKKKKKKKKKKKKKKKKKK KGGAP
      (K){20}
24: KKKK KKKKKKKKKKKKKKKKKKK GGAP

AA012098 ck: 8434 len: 39 1 Aa012098 Human polypeptide SEQ ID NO 251
      (R,K){20}
      (K){20}
13: TCKFG KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
14: CKFGK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
15: KFGKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
16: FGKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
17: GKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
18: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
```





1 AAO12203 ck: 4083 len: 41 i Aao12203 Human polypeptide SEQ ID NO 26095.  
(R,K){20}

6: SMCCL KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

7: WCCLK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

8: CCLKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

9: CLKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

1 AAO12215 ck: 6903 len: 73 i Aao12215 Human polypeptide SEQ ID NO 26107.  
(R,K){20}

50: PPPLP KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

51: PDLPK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

1 AAO12243 ck: 8474 len: 39 i Aao12243 Human polypeptide SEQ ID NO 26135.  
(R,K){20}

11: MISFI KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

12: ISFII KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

13: SPFKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

14: FIKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

15: IKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

1 AAO12250 ck: 1538 len: 69 i Aao12250 Human polypeptide SEQ ID NO 26142.  
(R,K){20}

31: SPSNL KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

32: PSNLK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

33: SNLKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

34: NLKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

35: LKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

36: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

37: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

1 AAO12274 ck: 2739 len: 37 i Aao12274 Human polypeptide SEQ ID NO 26136.  
(R,K){20}

38: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

39: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

40: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

41: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

42: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

43: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

44: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

45: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

14: KGLIN KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

15: GLINK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

16: LINKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

17: INKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

1 AAO12280 ck: 7481 len: 66 i Aao12280 Human polypeptide SEQ ID NO 26137.  
(R,K){20}

29: HPLPK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

30: IPLPK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

31: PPLPK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

32: PLKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

33: LKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

34: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

35: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

36: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

37: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

38: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

```
1
39: KKKKK (K){20} KKKSG
40: KKKKK (K){20} KKS GG
41: KKKKK (K){20} KSGGG
42: KKKKK (K){20} SGGGA
AA012447 ck: 5017 len: 47 1 Aa012447 Human polypeptide SEQ ID NO 26339.
(R,K){20}
(K){20}
14: HLVC E KKKKKKKKKKKKKKKKKKK KRGGA
(R){20}
15: LVCE K KKKKKKKKKKKKKKKKKKK RGAL
(R,K){20}
16: VCE K KKKKKKKKKKKKKKKKKKK GGALK
AA012476 ck: 837 len: 104 1 Aa012476 Human polypeptide SEQ ID NO 26369.
(R,K){20}
(K){20}
65: FCHQ KKKKKKKKKKKKKKKKKKK KGFEP
(K){20}
66: FCHQ KKKKKKKKKKKKKKKKKKK GFFEP
AA012548 ck: 6973 len: 74 1 Aa012548 Human polypeptide SEQ ID NO 26440.
(R,K){20}
(K){20}
38: AVLPL KKKKKKKKKKKKKKKKKKK KKG GG
(K){20}
39: VLPL KKKKKKKKKKKKKKKKKKK KGGV
(K){20}
40: LPLK KKKKKKKKKKKKKKKKKKK GGGVF
AA012553 ck: 1903 len: 33 1 Aa012553 Human polypeptide SEQ ID NO 26445.
(R,K){20}
(K){20}
9: SCCEI KKKKKKKKKKKKKKKKKKK KRGAP
(K){20}
10: CCEI KKKKKKKKKKKKKKKKKKK RGAP
(R,K){20}
11: CEI KKKKKKKKKKKKKKKKKKK GAP
AA013164 ck: 2798 len: 71 1 Aa013164 Human polypeptide SEQ ID NO 27056.
(R,K){20}
(R,K){20}
35: RPL KKKKKKKKKKKKKKKKKKK KEMFK
(K){20}
36: PL KKKKKKKKKKKKKKKKKKK EMFK
```

```
1
AA013785 ck: 4846 len: 99 1 Aa013785 Human polypeptide SEQ ID NO 274
(R,K){20}
(K){20}
53: TNNLI KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
54: NNL KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
55: NL KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
56: L KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
57: I KKKKKKKKKKKKKKKKKKK KKKG
(K){20}
58: K KKKKKKKKKKKKKKKKKKK KKG G
(K){20}
59: K KKKKKKKKKKKKKKKKKKK KGGP
(K){20}
60: K KKKKKKKKKKKKKKKKKKK GGGP
AA013785 ck: 6241 len: 100 1 Aa013785 Human polypeptide SEQ ID NO 276
(R,K){20}
(K){20}
41: LYAP KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
42: YAP KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
43: AP KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
44: P KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
45: PK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
46: K KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
47: K KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
48: K KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
49: K KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
50: K KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
51: K KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
52: K KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
53: K KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
54: K KKKKKKKKKKKKKKKKKKK KKKK
```

55: KKKKK (K){20} KKKKKKKKKKKKKKKKK KKKAGG  
56: KKKKK (K){20} KKKKKKKKKKKKKKKKK KAGGG  
57: KKKKK (K){20} KKKKKKKKKKKKKKKKK AGGGG  
AAU17983 ck: 7016 len: 315 1 Aau17983 Human immunoglobulin polypeptide S  
(R,K){20}  
(R,K){20}  
273: QVFAP KKKKKKKKKKKKKKKKK KGGRS  
(K){20}  
274: VFAPR KKKKKKKKKKKKKKKKK GGRSR  
AAU18049 ck: 7611 len: 315 1 Aau18049 Human immunoglobulin polypeptide S  
(R,K){20}  
(R,K){20}  
273: QVFAP KKKKKKKKKKKKKKKKK KGGRS  
(K){20}  
274: VFAPR KKKKKKKKKKKKKKKKK GGRSR  
AAU14119 ck: 5383 len: 86 1 Aam14119 Peptide #553 encoded by probe for  
(R,K){20}  
(R,K){20}  
15: RRRRG RRRRRKKKKKKKKKKKKKK KRRR  
(R,K){20}  
16: RRRGR RRRRRKKKKKKKKKKKKKK KRRR  
(R,K){20}  
17: RRGRR RRRRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
18: RGRRR RRRRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
19: GRRRR RRRRRKKKKKKKKKKKKKK RRRR  
(R,K){20}  
20: RRRRR KKKKKKKKKKKKKKKKKRR RRRR  
(R,K){20}  
21: RRRRR KKKKKKKKKKKKKKKKKRR RRRR  
(R,K){20}  
22: RRRRK KKKKKKKKKKKKKKKKKRR RRRR  
(R,K){20}  
23: RRRKK KKKKKKKKKKKKKKKKKRR RRRR  
(R,K){20}  
24: RRRKK KKKKKKKKKKKKKKKKKRR RRRR  
(R,K){20}  
25: KKKKK KKKKKKKKKKKKKKKKKRR RRRR  
(R,K){20}  
26: KKKKK KKKKKKKKKKKKKKKKKRR RRRR  
(R,K){20}  
27: KKKKK KKKKKKKKKKKKKKKKKRR RRRR  
(R,K){20}

28: KKKKK KKKKKKKKKKKKKKKKKRR RRRR  
(R,K){20}  
29: KRRKK KKKKKKKKKKKKKKKKKRR RRRR  
(R,K){20}  
30: KRRKK KKKKKKKKKKKKKKKKKRR RRRR  
(R,K){20}  
31: RKKKK KKKKKKKKKKKKKKKKKRR RRRR  
(R,K){20}  
32: KKKKK KKKKKKKKKKKKKKKKKRR RRRR  
(R,K){20}  
33: KKKKK KKKKKKKKKKKKKKKKKRR RRRR  
(R,K){20}  
34: KKKKK KKKKKKKKKKKKKKKKKRR RRRR  
(R,K){20}  
35: KKKKK KKKKKKKKKKKKKKKKKRR RRRR  
(R,K){20}  
36: KKKKK KRRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
37: RKKKK RRRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
38: KKKKK RRRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
39: KKKKK RRRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
40: KKKKK RRRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
41: KRRRR RRRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
42: RRRRR RRRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
43: RRRRR RRRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
44: RRRRR RRRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
45: RRRRR RRRRRRRRRRRRRRRRRRR RRRR  
(R){20}  
AAM14961 ck: 1334 len: 86 1 Aam14961 Peptide #1395 encoded by probe  
(R,K){20}  
(R,K){20}  
57: EEEEG RRRKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
58: EEEGR RRRKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
59: EEGRR RRRKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
60: EGRRR KKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}  
61: GRRRR KKKKKKKKKKKKKKKKK KKKKK  
(R,K){20}

62: RRRK KRRKKKKKKKKKKKKKK  
(R,K){20}  
63: RRRK KRRKKKKKKKKKKKKKK  
(R,K){20}  
64: RRRK KRRKKKKKKKKKKKKKK  
(R,K){20}  
65: RRRK KRRKKKKKKKKKKKKKK  
(R,K){20}  
66: RRRK KRRKKKKKKKKKKKKKK  
(R,K){20}  
67: RRRK KRRKKKKKKKKKKKKKK  
(R,K){20}

AA15038 ck: 9082 len: 167 ! AA15038 Peptide #1472 encoded by probe for

33: EERG RRRRRRRRRRRRRRRRR  
(R,K){20}  
34: EERG RRRRRRRRRRRRRRRRR  
(R,K){20}  
35: GRRR RRRRRRRRRRRRRRRRR  
(R,K){20}

AA15527 ck: 2276 len: 89 ! AA15527 Peptide #1961 encoded by probe for

23: EEEE KRRKKKKKKKKKKKKKK  
(R,K){20}  
46: KKEE KRRKKKKKKKKKKKKKK  
(R,K){20}  
47: KKEE KRRKKKKKKKKKKKKKK  
(R,K){20}  
48: EEEK KRRKKKKKKKKKKKKKK  
(R,K){20}  
49: EEEK KRRKKKKKKKKKKKKKK  
(R,K){20}

AA15826 ck: 1939 len: 130 ! AA15826 Peptide #2260 encoded by probe for

42: EGRE RRRRRRRRRRRRRRRRR  
(R,K){20}  
43: GRER RRRRRRRRRRRRRRRRR  
(R,K){20}  
44: RKEE RRRRRRRRRRRRRRRRR  
(R,K){20}  
45: KEER RRRRRRRRRRRRRRRRR  
(R,K){20}  
46: EERR RRRRRRRRRRRRRRRRR  
(R,K){20}  
47: RRRR RRRRRRRRRRRRRRRRR  
(R,K){20}  
48: RRRR RRRRRRRRRRRRRRRRR  
(R,K){20}

49: RRRR RRRRRRRRRRRRRRRRR  
(R,K){20}  
50: RRRR RRRRRRRRRRRRRRRRR  
(R,K){20}  
51: RRRR RRRRRRRRRRRRRRRRR  
(R,K){20}  
52: RRRR RRRRRRRRRRRRRRRRR  
(R,K){20}  
53: RRRR RRRRRRRRRRRRRRRRR  
(R,K){20}  
54: RRRR RRRRRRRRRRRRRRRRR  
(R,K){20}

AA16123 ck: 3607 len: 88 ! AA16123 Peptide #2557 encoded by probe

39: EKRK KRRKKKKKKKKKKKKKK  
(R,K){20}  
40: RKRK KRRKKKKKKKKKKKKKK  
(R,K){20}  
41: KREK KRRKKKKKKKKKKKKKK  
(R,K){20}  
42: REKK KRRKKKKKKKKKKKKKK  
(R,K){20}  
43: EKKK KRRKKKKKKKKKKKKKK  
(R,K){20}  
44: KKKK KRRKKKKKKKKKKKKKK  
(R,K){20}  
45: KKKK KRRKKKKKKKKKKKKKK  
(R,K){20}  
46: KKKK KRRKKKKKKKKKKKKKK  
(R,K){20}  
47: KKKK KRRKKKKKKKKKKKKKK  
(R,K){20}  
48: KKKK KRRKKKKKKKKKKKKKK  
(R,K){20}  
49: KKKK KRRKKKKKKKKKKKKKK  
(R,K){20}  
50: KKKK KRRKKKKKKKKKKKKKK  
(R,K){20}  
51: KKKK KRRKKKKKKKKKKKKKK  
(R,K){20}  
52: KKKK KRRKKKKKKKKKKKKKK  
(R,K){20}  
53: KKKK KRRKKKKKKKKKKKKKK  
(R,K){20}  
54: KKKK KRRKKKKKKKKKKKKKK  
(R,K){20}  
55: KKKK KRRKKKKKKKKKKKKKK  
(R,K){20}

```

56: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
57: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
58: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
59: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
60: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
61: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
62: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
63: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
64: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
65: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
66: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
67: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
68: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
69: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

AAM1691 ck: 3937 len: 85 1 Aam1691 Peptide #3125 encoded by probe for
1: (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
2: K (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
3: KK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
4: KKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
5: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
6: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
7: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
8: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
9: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

```

```

10: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
11: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK EEEEX

AAM17388 ck: 2686 len: 71 1 Aam17388 Peptide #3822 encoded by probe
20: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
21: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
22: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
23: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
24: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
25: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
26: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
27: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
28: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
29: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
30: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
31: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
32: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
33: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
34: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
35: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
36: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
37: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
38: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
39: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
40: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK

```

1

```
41: KKKKK (R,K){20} KKKKK
42: KKKKK (R,K){20} KKKKK
43: KKKKK (R,K){20} KKKKK
44: KKKKK (R,K){20} KKKKK
45: KKKKK (R,K){20} KKKKK
46: KKKKK (R,K){20} KKKSA
47: KKKKK (K){20} KKSAA
48: KKKKK (K){20} KKSAA
49: KKKKK (K){20} SAH
AAM18532 ck: 1560 len: 88 ! Aam18532 Peptide #4966 encoded by probe for
43: RRRRG (R,K){20} RRRRR
44: ERGGG (R){20} RRRRR
45: RRGRR (R){20} RRRRR
46: RGGRR (R){20} RRRRR
47: GRRRR (R){20} RRRRR
48: RRRRR (R){20} RRRRR
49: RRRRR (R){20} RRRRR
50: RRRRR (R){20} RRRRR
51: RRRRR (R){20} RRRRR
52: RRRRR (R){20} RRRRR
53: RRRRR (R){20} RRRRR
54: RRRRR (R){20} RRRRR
55: RRRRR (R){20} RRRRR
56: RRRRR (R){20} RRRRR
```

1

1

1

```
57: RRRRR (R){20} RRRRR
58: RRRRR (R){20} RRRRR
59: RRRRR (R){20} RRRRR
60: RRRRR (R){20} RRRRR
61: RRRRR (R){20} RRRRR
62: RRRRR (R){20} RRRRR
AAM20089 ck: 8343 len: 66 ! Aam20089 Peptide #6523 encoded by probe
6: ETERE (R,K){20} KKKKK
7: TERER (R,K){20} KKKKK
8: EREKK (R,K){20} KKKKK
9: REKKK (R,K){20} NKKKK
AAM21941 ck: 4895 len: 51 ! Aam21941 Peptide #8375 encoded by probe
18: LERPM (R,K){20} KKKKK
19: FKPMR (R,K){20} KKKKK
20: KPMRK (R,K){20} LTTT
AAM26526 ck: 5383 len: 86 ! Aam26526 Peptide #563 encoded by probe
15: RRRRG (R,K){20} KKKKK
16: RRRGR (R,K){20} KKKKK
17: RGGRR (R,K){20} KKKKK
18: RGGRR (R,K){20} KKKKK
19: GRRRR (R,K){20} KKKKK
20: RRRRR (R,K){20} KKKKK
21: RRRRK (R,K){20} KKKKK
```

22: RRRKK (R,K){20} KKKKKKKKKKKKKRRRR RRRR  
23: RRRKK (R,K){20} KKKKKKKKKKKKKRRRR RRRR  
24: RKKKK (R,K){20} KKKKKKKKKKKKKRRRR RRRR  
25: KKKKK (R,K){20} KKKKKKKKKKKKKRRRR RRRR  
26: KKKKK (R,K){20} KKKKKKKKKKKKKRRRR RRRR  
27: KKKKK (R,K){20} KKKKKKKKKKKKKRRRR RRRR  
28: KKKKK (R,K){20} KKKKKKKKKKKKKRRRR RRRR  
29: KRRKK (R,K){20} KKKKKKKKKKKKKRRRR RRRR  
30: KRRKK (R,K){20} KKKKKKKKKKKKKRRRR RRRR  
31: KKKKK (R,K){20} KKKKKKKKKKKKKRRRR RRRR  
32: KKKKK (R,K){20} KKKKKKKKKKKKKRRRR RRRR  
33: KKKKK (R,K){20} KKKKKKKKKKKKKRRRR RRRR  
34: KKKKK (R,K){20} KKKKKKKKKKKKKRRRR RRRR  
35: KRRKK (R,K){20} KKKKKKKKKKKKKRRRR RRRR  
36: KRRKK (R,K){20} KRRRRRRRRRRRRRRRR RRRR  
37: KRRKK (R,K){20} RRRRRRRRRRRRRRRRR RRRR  
38: KKKKK (R,K){20} RRRRRRRRRRRRRRRRR RRRR  
39: KKKKK (R,K){20} RRRRRRRRRRRRRRRRR RRRR  
40: KRRRR (R,K){20} RRRRRRRRRRRRRRRRR RRRR  
41: KRRRR (R,K){20} RRRRRRRRRRRRRRRRR RRRR  
42: RRRRR (R,K){20} RRRRRRRRRRRRRRRRR RRRR  
43: RRRRR (R,K){20} RRRRRRRRRRRRRRRRR RRRR  
44: RRRRR (R,K){20} RRRRRRRRRRRRRRRRR RRRR  
45: RRRRR (R,K){20} RRRRRRRRRRRRRRRRR RRRR

1

AAM27395 ck: 1334 len: 86 ! Aam27395 Peptide #1432 encoded by probe

57: EEEEG (R,K){20} RRRRRRRRRRRRRRRRRR KKKKK  
58: EEEGR (R,K){20} RRRRRRRRRRRRRRRRRR KKKKK  
59: EEEGR (R,K){20} RRRRRRRRRRRRRRRRRR KKKKK  
60: EEEGR (R,K){20} RRRRRRRRRRRRRRRRRR KKKKK  
61: GRRRK (R,K){20} KKKKKKKKKKKKKKKKK KKKKK  
62: RRRKK (R,K){20} KKKKKKKKKKKKKKKKK KKKKK  
63: RRRKK (R,K){20} KKKKKKKKKKKKKKKKK KKKKK  
64: RRRKK (R,K){20} KKKKKKKKKKKKKKKKK KKKKK  
65: KRRKK (R,K){20} KKKKKKKKKKKKKKKKK KKKKK  
66: KRRKK (R,K){20} KKKKKKKKKKKKKKKKK KKKKK  
67: KRRKK (R,K){20} KKKKKKKKKKKKKKKKK KKKKK

1

AAM27479 ck: 9082 len: 167 ! Aam27479 Peptide #1516 encoded by probe

33: EEEGR (R,K){20} RRRRRRRRRRRRRRRRRR RRRRR  
34: EEEGR (R,K){20} RRRRRRRRRRRRRRRRRR RRRRR  
35: GGGRR (R,K){20} RRRRRRRRRRRRRRRRRR RRRRR

1

AAM28016 ck: 2276 len: 89 ! Aam28016 Peptide #2053 encoded by probe

23: EEEEE (R,K){20} KKKKKKKKKKKKKKKKK EEEEE  
46: KKEEE (R,K){20} KKKKKKKKKKKKKKKKK KKEEE  
47: KEEER (R,K){20} KKKKKKKKKKKKKKKKK KKEEE  
48: EEEKK (R,K){20} KKKKKKKKKKKKKKKKK KKEEE  
49: EEEKK (R,K){20} KKKKKKKKKKKKKKKKK EEEEE

1

AAM28336 ck: 1939 len: 130 ! Aam28336 Peptide #2373 encoded by probe

```
42: EGKKE RRRRRRRRRRRRRRRRRRRR RRRRR
    (R,K){20}
43: GKRRR RRRRRRRRRRRRRRRRRRRR RRRRR
    (R,K){20}
44: RKERR RRRRRRRRRRRRRRRRRRRR RRRRR
    (R,K){20}
45: KEERRR RRRRRRRRRRRRRRRRRRRR RRRRR
    (R,K){20}
46: EERRR RRRRRRRRRRRRRRRRRRRR RRRRR
    (R,K){20}
47: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
    (R,K){20}
48: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
    (R,K){20}
49: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
    (R,K){20}
50: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
    (R,K){20}
51: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
    (R,K){20}
52: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
    (R,K){20}
53: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
    (R,K){20}
54: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
    (R,K){20}

AAM28616 ck: 3607 len: 88 1 Aam28616 Peptide #2653 encoded by probe for
    (R,K){20}
39: ERKRE RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
40: RKREK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
41: KREKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
42: REKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
43: EKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
44: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
45: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
46: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
47: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
48: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
```

```
49: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
50: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
51: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
52: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
53: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
54: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
55: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
56: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
57: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
58: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
59: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
60: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
61: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
62: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
63: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
64: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
65: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
66: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
67: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
68: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
69: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}

AAM29178 ck: 3937 len: 85 1 Aam29178 Peptide #3215 encoded by probe
    (R,K){20}
1: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
2: KKKKK RRRRRRRRRRRRRRRRRRRR RRRRR
    (K){20}
```



3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
AAM30989 ck: 1560 len: 88 1 Aam30989 Peptide #5026 encoded by probe for  
(R,K){20}  
43: REERG RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}  
44: ERGRG RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}  
45: RGRGR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}  
46: RGRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}  
47: GRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}  
48: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}  
49: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}  
50: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}  
51: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}  
52: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}  
53: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}  
54: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}  
55: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}  
56: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}

57: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}  
58: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}  
59: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}  
60: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}  
61: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}  
62: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}  
AAM31173 ck: 2324 len: 36 1 Aam31173 Peptide #5210 encoded by probe  
(R,K){20}  
16: KERKT KKKRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}  
17: ERKTK KKKRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}

AAM33953 ck: 8343 len: 66 1 Aam33953 Peptide #7990 encoded by probe  
(R,K){20}

6: ETERE KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(R,K){20}

7: TEREX KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(R,K){20}

8: EREKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(R,K){20}

9: REKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(R,K){20}

AAM36456 ck: 2394 len: 57 1 Aam36456 Peptide #10493 encoded by probe  
(R,K){20}

20: EEEEG RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}

21: EEERG RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}

22: EEGRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}

23: EGRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}

24: GRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}

25: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}

AAM37018 ck: 4228 len: 24 1 Aam37018 Peptide #11055 encoded by probe  
(R,K){20}

1: RRRRRRKKKKRRRRRRRRRRRRRRRRRRRRRRRR  
(R,K){20}

1

```
2: R RRRRRRRRRRRRRRRRRRRR KRT
   (R,K){20}
3: RR RRRRRRRRRRRRRRRRRR RT
   (R,K){20}
4: RRR RRRRRRRRRRRRRRRRRR T
   (R,K){20}

AAM37794 ck: 3301 len: 52 ! Aam37794 Peptide #11831 encoded by probe fd
   (R,K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
13: KKKKK (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
14: KKKKK (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
15: KKKKK (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
16: KKKKK (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
17: KKKKK (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
18: KKKKK (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
19: KKKKK (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
20: KKKKK (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
21: KKKKK (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
22: KKKKK (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
23: KKKKK (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
24: KKKKK (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
25: KKKKK (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
26: KKKKK (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
27: KKKKK (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
28: KKKKK (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
29: KKKKK (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
30: KKKKK (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
31: KKKKK (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
```

1

1

```
AAM38273 ck: 4895 len: 51 ! Aam38273 Peptide #12310 encoded by probe
   (R,K){20}
18: LKPKK KKKKKKKKKKKKKKKKKRR KKLTT
   (R,K){20}
19: KKKKK KKKKKKKKKKKKKKKKKRR KKLTT
   (R,K){20}
20: KKKKK KKKKKKKKKKKKKKKKKRR KKLTT
   (R,K){20}

AAU04283 ck: 8137 len: 45 ! Aau04283 Trimeric fusogenic peptide #2 u
   (R,K){20}
4: YKA KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
5: YKA KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
6: YKA KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
```

```

1
Au04285 ck: 4361 len: 59 i Au04285 Nuclear ligand #2 used in nucleic
(R,K){20}
(K){20}
18: ADYKA KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
19: PYKAK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
20: YKAKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
21: KAKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
22: AKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}

```

```

1
AAU04287 ck: 4925 len: 100 ! AAU04287 Poly-L-Lysine used in nucleic

1: (R,K){20}
   KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

2: K KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

3: KK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

4: KKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

```

24: KKKK (K){20}  
 25: KKKK (K){20}  
 26: KKKK (K){20}  
 27: KKKK (K){20}  
 28: KKKK (K){20}  
 29: KKKK (K){20}  
 30: KKKK (K){20}  
 31: KKKK (K){20}  
 32: KKKK (K){20}  
 33: KKKK (K){20}  
 34: KKKK (K){20}  
 35: KKKK (K){20}  
 36: KKKK (K){20}  
 37: KKKK (K){20}  
 38: KKKK (K){20}  
 39: KKKK (K){20}  
 40: KKKK (K){20}  
 41: KKKK (K){20}  
 42: KKKK (K){20}  
 43: KKKK (K){20}  
 44: KKKK (K){20}  
 45: KKKK (K){20}  
 46: KKKK (K){20}  
 47: KKKK (K){20}

48: KKKK (K){20}  
 49: KKKK (K){20}  
 50: KKKK (K){20}  
 51: KKKK (K){20}  
 52: KKKK (K){20}  
 53: KKKK (K){20}  
 54: KKKK (K){20}  
 55: KKKK (K){20}  
 56: KKKK (K){20}  
 57: KKKK (K){20}  
 58: KKKK (K){20}  
 59: KKKK (K){20}  
 60: KKKK (K){20}  
 61: KKKK (K){20}  
 62: KKKK (K){20}  
 63: KKKK (K){20}  
 64: KKKK (K){20}  
 65: KKKK (K){20}  
 66: KKKK (K){20}  
 67: KKKK (K){20}  
 68: KKKK (K){20}  
 69: KKKK (K){20}  
 70: KKKK (K){20}  
 71: KKKK (K){20}  
 72: KKKK (K){20}



(R,K){20}  
63: RRRR RRRRRRRRRRRRRRRR KKKK

(R,K){20}  
64: RRRR RRRRRRRRRRRRRRRR KKK

(R,K){20}  
65: KRRR RRRRRRRRRRRRRRRR KK

(R,K){20}  
66: KRRR RRRRRRRRRRRRRRRR K

(K){20}  
67: KRRR RRRRRRRRRRRRRRRR

AA02768 ck: 9082 len: 167 ! Aa02768 Peptide #1450 encoded by probe for

(R,K){20}  
33: EGRG RRRRRRRRRRRRRRRR RGGR

(R,K){20}  
34: EGRG RRRRRRRRRRRRRRRR RGGR

(R,K){20}  
35: GRGR RRRRRRRRRRRRRRRR GGGR

AA03278 ck: 2276 len: 89 ! Aa03278 Peptide #1960 encoded by probe for

(R,K){20}  
23: EEEE RRRRRRRRRRRRRRRR EEEK

(K){20}  
46: KEEE RRRRRRRRRRRRRRRR KKEE

(K){20}  
47: KEEE RRRRRRRRRRRRRRRR KKEE

(R,K){20}  
48: EEEK RRRRRRRRRRRRRRRR EEEE

(R,K){20}  
49: EEEK RRRRRRRRRRRRRRRR EEEE

AA03564 ck: 1939 len: 130 ! Aa03564 Peptide #2246 encoded by probe for

(R,K){20}  
42: EGKE RRRRRRRRRRRRRRRR RRRR

(R,K){20}  
43: GKER RRRRRRRRRRRRRRRR RRRR

(R,K){20}  
44: KERR RRRRRRRRRRRRRRRR RRRR

(R,K){20}  
45: KERR RRRRRRRRRRRRRRRR RRRR

(R,K){20}  
46: EERR RRRRRRRRRRRRRRRR RRRR

(R,K){20}  
47: RRRR RRRRRRRRRRRRRRRR RRRR

(R,K){20}  
48: RRRR RRRRRRRRRRRRRRRR RRRR

(R,K){20}  
49: RRRR RRRRRRRRRRRRRRRR KKKK

(R,K){20}  
50: RRRR RRRRRRRRRRRRRRRR KKKK

(R,K){20}  
51: RRRR RRRRRRRRRRRRRRRR KKKK

(R,K){20}  
52: RRRR RRRRRRRRRRRRRRRR KKKK

(R,K){20}  
53: RRRR RRRRRRRRRRRRRRRR KKKK

(R,K){20}  
54: RRRR RRRRRRRRRRRRRRRR EEEE

AA03848 ck: 3607 len: 88 ! Aa03848 Peptide #2530 encoded by probe

(R,K){20}  
39: EKRK RRRRRRRRRRRRRRRR KKKK

(K){20}  
40: KRER RRRRRRRRRRRRRRRR KKKK

(K){20}  
41: KREK RRRRRRRRRRRRRRRR KKKK

(K){20}  
42: REKK RRRRRRRRRRRRRRRR KKKK

(K){20}  
43: EKRR RRRRRRRRRRRRRRRR KKKK

(K){20}  
44: KRRR RRRRRRRRRRRRRRRR KKKK

(K){20}  
45: KRRR RRRRRRRRRRRRRRRR KKKK

(K){20}  
46: KRRR RRRRRRRRRRRRRRRR KKKK

(K){20}  
47: KRRR RRRRRRRRRRRRRRRR KKKK

(K){20}  
48: KRRR RRRRRRRRRRRRRRRR KKKK

(K){20}  
49: KRRR RRRRRRRRRRRRRRRR KKKK

(K){20}  
50: KRRR RRRRRRRRRRRRRRRR KKKK

(K){20}  
51: KRRR RRRRRRRRRRRRRRRR KKKK

(K){20}  
52: KRRR RRRRRRRRRRRRRRRR KKKK

(K){20}  
53: KRRR RRRRRRRRRRRRRRRR KKKK

(K){20}  
54: KRRR RRRRRRRRRRRRRRRR KKKK

(K){20}  
55: KRRR RRRRRRRRRRRRRRRR KKKK

(K){20}  
56: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
57: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
58: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
59: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
60: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
61: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
62: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
63: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
64: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
65: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
66: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
67: KKKKK KKKKKKKKKKKKKKKKKKK KK  
(K){20}  
68: KKKKK KKKKKKKKKKKKKKKKKKK K  
(K){20}  
69: KKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}  
AAM04408 ck: 3937 len: 85 1 Aam04408 Peptide #3090 encoded by probe for  
(R,K){20}  
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
2: K KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK  
(K){20}

(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKKKK KKEEE  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKK EEEEE  
AAM06100 ck: 1560 len: 88 1 Aam06100 Peptide #4782 encoded by probe  
(R,K){20}  
43: RRRRG RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
44: ERGRG RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
45: RGRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
46: RGRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
47: GRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
48: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
49: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
50: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
51: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
52: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
53: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
54: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
55: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
56: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
57: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
58: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
59: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
60: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
61: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
62: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR  
(R){20}  
AAG73687 ck: 3063 len: 25 1 Aag73687 Human colon cancer antigen prot

```
1
  8: MDDTF KKKKKKKKKKKKKKKKKKKKK XX
    (K){20}
  9: MMTFK KKKKKKKKKKKKKKKKKKKKK X
    (K){20}
AAG73729 ck: 783 len: 83 1 Aag73729 Human colon cancer antigen protein
  (R,K){20}
  49: LGPCE KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  50: GPCEK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  51: PCEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  52: CEKKK KKKKKKKKKKKKKKKKKKKKK KKKKG
    (K){20}
  53: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKG
    (K){20}
  54: KKKKK KKKKKKKKKKKKKKKKKKKKK KKGGR
    (K){20}
  55: KKKKK KKKKKKKKKKKKKKKKKKKKK KGGFK
    (K){20}
  56: KKKKK KKKKKKKKKKKKKKKKKKKKK GGRFK
    (K){20}
AAG73810 ck: 3374 len: 88 1 Aag73810 Human colon cancer antigen protein
  (R,K){20}
  44: FGOTX KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  45: GQTXK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  46: QTXKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  47: TXKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  48: XKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  49: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  50: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  51: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  52: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  53: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  54: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}

1
  55: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  56: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  57: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  58: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  59: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  60: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  61: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKG
    (K){20}
  62: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKG
    (K){20}
  63: KKKKK KKKKKKKKKKKKKKKKKKKKK KKGGP
    (K){20}
  64: KKKKK KKKKKKKKKKKKKKKKKKKKK KGGFK
    (K){20}
  65: KKKKK KKKKKKKKKKKKKKKKKKKKK GGPFK
    (K){20}
AAG73895 ck: 1887 len: 43 1 Aag73895 Human colon cancer antigen prot
  (R,K){20}
  18: VRPRV RKKKKKKKKKKKKKKKKKKKK KKKKG
    (R,K){20}
  19: RPRVR KKKKKKKKKKKKKKKKKKKKK KKKKG
    (K){20}
  20: PRVRK KKKKKKKKKKKKKKKKKKKKK KKGK
    (K){20}
  21: RVRKK KKKKKKKKKKKKKKKKKKKKK KKG
    (K){20}
  22: VRKKK KKKKKKKKKKKKKKKKKKKKK GG
    (K){20}
AAG74218 ck: 8659 len: 104 1 Aag74218 Human colon cancer antigen prot
  (R,K){20}
  75: PLGGQ KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  76: LGGQK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
  77: GGQKK KKKKKKKKKKKKKKKKKKKKK KKKKG
    (K){20}
  78: GQKKK KKKKKKKKKKKKKKKKKKKKK KKKKG
    (K){20}
  79: QKKKK KKKKKKKKKKKKKKKKKKKKK KKGKP
    (K){20}
  80: KKKKK KKKKKKKKKKKKKKKKKKKKK XGKPF
    (K){20}
```



AAg74527 ck: 2664 len: 40 : Aag74527 Human colon cancer antigen protein

1 (R,K){20}

(K){20}

9: CULLY KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

10: LLLY KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

11: LLYK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

12: LYKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

13: YKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

14: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

15: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

16: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

17: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

AAg74650 ck: 1596 len: 69 : Aag74650 Human colon cancer antigen protein

1 (R,K){20}

(K){20}

36: LQCRQ KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

37: QCRQK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

38: CRQKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

39: RQKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

40: QKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

41: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

42: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

43: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

44: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

45: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

46: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

AAg74793 ck: 8497 len: 152 : Aag74793 Human colon cancer antigen protein

(R,K){20}

(K){20}

122: SHTQ KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

123: SHTQ KKKKKKKKKKKKKKKKKKKKKKKKK

AAg74907 ck: 1215 len: 98 : Aag74907 Human colon cancer antigen protein

1

(R,K){20}

(K){20}

57: NLRE KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

58: LRKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

59: RKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

60: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

61: EKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

62: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

63: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

64: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

65: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

66: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

67: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

68: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

69: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

70: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

71: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

72: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

73: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

74: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

75: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

76: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

AAg75215 ck: 3913 len: 155 : Aag75215 Human colon cancer antigen protein

(R,K){20}

135: RSSAP KKKKKKKKKKKKKKKKKKK K  
(K){20}  
136: SSAPK KKKKKKKKKKKKKKKKKKK  
(K){20}

AAG75886 ck: 4235 len: 71 1 Aag75886 Human colon cancer antigen protein  
(R,K){20}  
47: KKKKK KKKKKKKKKKKKKKKKKKK KXGX  
(K){20}  
48: KKKKK KKKKKKKKKKKKKKKKKKK XGX  
(K){20}

AAE01796 ck: 4416 len: 72 1 Aae01796 Human gene 27 encoded secreted pro  
(R,K){20}  
47: LPTEL KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
48: PTELK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
49: TELKK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
50: FLKKK KKKKKKKKKKKKKKKKKKK KKK  
(K){20}  
51: LKKKK KKKKKKKKKKKKKKKKKKK KK  
(K){20}  
52: KKKKK KKKKKKKKKKKKKKKKKKK K  
(K){20}  
53: KKKKK KKKKKKKKKKKKKKKKKKK  
(K){20}

AAE01848 ck: 5584 len: 73 1 Aae01848 Human gene 27 encoded secreted pro  
(R,K){20}  
47: LPTEL KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
48: PTELK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
49: TELKK KKKKKKKKKKKKKKKKKKK KKKKI  
(K){20}  
50: FLKKK KKKKKKKKKKKKKKKKKKK KKKI  
(K){20}  
51: LKKKK KKKKKKKKKKKKKKKKKKK KKI  
(K){20}  
52: KKKKK KKKKKKKKKKKKKKKKKKK KI  
(K){20}  
53: KKKKK KKKKKKKKKKKKKKKKKKK I  
(K){20}

AAB90574 ck: 1431 len: 530 1 Aab90574 Human secreted protein, SEQ ID NO:  
(R,K){20}  
511: LHAPP KKKKKKKKKKKKKKKKKKKK  
(K){20}

1  
AAB45846 ck: 8137 len: 45 1 Aab45846 Nucleic acid transporter system  
(R,K){20}  
4: YKA KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
5: YKAK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
6: YKAK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
7: KAKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
8: AKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}  
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

AAB45848 ck: 4361 len: 59 1 Aab45848 Nucleic acid transporter system  
(R,K){20}  
18: APYKA KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

19: PYAK (K){20} KKKK  
(K){20}  
20: YKAK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
21: KAKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
22: AKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
23: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
24: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
25: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
26: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
27: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
28: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
29: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
30: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
31: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
32: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
33: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
34: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
35: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
36: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
37: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
38: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

AAb45850 ck: 4925 len: 100 1 Aab45850 Nucleic acid transporter system pe

1: (R,K){20} KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
2: K (K){20} KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
3: KK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

4: KK (K){20} KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
5: KK KKKKKKKKKKKKKKKKKKK KKKK  
(K){20}  
6: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
7: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
8: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
9: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
10: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
11: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
12: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
13: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
14: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
15: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
16: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
17: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
18: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
19: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
20: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
21: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
22: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
23: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
24: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
25: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
26: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}  
27: KKKK KKKKKKKKKKKKKKKKK KKKK  
(K){20}

28: KKKK (K){20} KKKK  
29: KKKK (K){20} KKKK  
30: KKKK (K){20} KKKK  
31: KKKK (K){20} KKKK  
32: KKKK (K){20} KKKK  
33: KKKK (K){20} KKKK  
34: KKKK (K){20} KKKK  
35: KKKK (K){20} KKKK  
36: KKKK (K){20} KKKK  
37: KKKK (K){20} KKKK  
38: KKKK (K){20} KKKK  
39: KKKK (K){20} KKKK  
40: KKKK (K){20} KKKK  
41: KKKK (K){20} KKKK  
42: KKKK (K){20} KKKK  
43: KKKK (K){20} KKKK  
44: KKKK (K){20} KKKK  
45: KKKK (K){20} KKKK  
46: KKKK (K){20} KKKK  
47: KKKK (K){20} KKKK  
48: KKKK (K){20} KKKK  
49: KKKK (K){20} KKKK  
50: KKKK (K){20} KKKK  
51: KKKK (K){20} KKKK

52: KKKK (K){20} KKKK  
53: KKKK (K){20} KKKK  
54: KKKK (K){20} KKKK  
55: KKKK (K){20} KKKK  
56: KKKK (K){20} KKKK  
57: KKKK (K){20} KKKK  
58: KKKK (K){20} KKKK  
59: KKKK (K){20} KKKK  
60: KKKK (K){20} KKKK  
61: KKKK (K){20} KKKK  
62: KKKK (K){20} KKKK  
63: KKKK (K){20} KKKK  
64: KKKK (K){20} KKKK  
65: KKKK (K){20} KKKK  
66: KKKK (K){20} KKKK  
67: KKKK (K){20} KKKK  
68: KKKK (K){20} KKKK  
69: KKKK (K){20} KKKK  
70: KKKK (K){20} KKKK  
71: KKKK (K){20} KKKK  
72: KKKK (K){20} KKKK  
73: KKKK (K){20} KKKK  
74: KKKK (K){20} KKKK  
75: KKKK (K){20} KKKK  
76: KKKK (K){20} KKKK

77: KKKKK (K){20}  
78: KKKKK (K){20}  
79: KKKKK (K){20}  
80: KKKKK (K){20}  
81: KKKKK (K){20}

AAB50247 ck: 8085 len: 154 1 Aab50247 Human breast cancer associated B72  
(R,K){20}

114: TOLRQ KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

115: QLFQR KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

116: LROK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

117: RQKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

118: OKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

119: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

120: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

121: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

122: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

123: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

124: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

125: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

126: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

127: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

128: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

ABB44830 ck: 5275 len: 38 1 ABB44830 Human protective sequence CNI-0074  
(R,K){20}

4: MAQ KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

5: MAQ KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

6: MAQR KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

7: AQKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

8: QKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

AAU69690 ck: 875 len: 46 1 AAU69690 Cell death protective sequence  
(R,K){20}

15: REKKS KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

16: EKSK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

17: KSKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

18: KSKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

19: SKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK  
(K){20}

1

25: KKKKK KKKKKKKKKKKKKKKKKKKKK GG  
AAU69736 ck: 5862 len: 50 ! Aau69736 Cell death protective sequence CMI

(R,K){20}  
(R,K){20}

11: ESALG KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

12: SALGR KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

13: ALGRK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

14: LGRKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

15: GRKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

16: RKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

19: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

20: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

21: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

22: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

23: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

24: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

25: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

26: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

27: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

28: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

29: KKKKK KKKKKKKKKKKKKKKKKKKKK GR

Databases searched:

EMBL, Release 7.0, Released on 28Apr2002, Formatted on 2May2002

Total finds: 7,093  
Total length: 114,001,827  
Total sequences: 766,495  
CPU time: 12:26.84